Inhibitor Synthesis and Biophysical Characterization of Protein–Ligand–Solvent Interactions

An Analysis of the Thermodynamics and Kinetics of Ligand Binding to Thermolysin

Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

dem
Fachbereich Pharmazie der
Philipps-Universität Marburg
vorgelegt

von
Pharmazeut / M.Sc. Med. Chem.
Jonathan Cramer
aus
Gießen

Marburg/Lahn 2017

Erstgutachter	Prof. Dr. Gerhard Klebe Institut für Pharmazeutische Chemie Philipps-Universität Marburg
Zweitgutachter	Prof. Dr. Martin Schlitzer Institut für Pharmazeutische Chemie Philipps-Universität Marburg

Eingereicht am 24.04.2017

Tag der Mündlichen Prüfung am 09.06.2017

Hochschulkennziffer: 1180

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. Gerhard Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von Mai 2013 bis März 2017 durchgeführt.

Erklärung zum Eigenanteil

Titel der Publikation	Autoren	geschätzter Eigenanteil	Status
An allyl protection and improved purification strategy enables the synthesis of functionalized phosphonamidate peptides *Synthesis* **2017**, 49: 1857–1866	Cramer, J., Klebe, G.	90%	Publiziert
Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein Bound Ligands *J. Med. Chem.* **2016**, 59: 10530–10548	Krimmer, S.G.,* Cramer, J.,* Betz, M., Fridh, V., Karlsson, R., Heine, A., Klebe, G.	30%	Publiziert
Paying the Price of Desolvation in Solvent-Exposed Protein Pockets: Impact of Solubilizing Polar Groups on Affinity and Binding Thermodynamics in a Series of Thermolysin Inhibitors *Manuscript submitted*	Cramer, J., Krimmer, S.G., Heine, A., Klebe, G.	70%	Eingereicht
Elucidating the Origin of Long Residence Time Binding for Inhibitors of the Metalloprotease Thermolysin *ACS Chem. Biol.* **2017**, 12: 225–233	Cramer, J.,* Krimmer, S.G.,* Fridh, V.,* Wulsdorf, T., Karlsson, R., Heine, A., Klebe, G.	20%	Publiziert
Watch out for the Red Herring: Surprising Reactivity of a Fragment Results in Biological Activity *Angew. Chem. Int. Ed.* **2017**, 7: 1908–1913 *Angew. Chem.* **2017**, 6: 1934–1940	Cramer, J.,* Schiebel, J.,* Wulsdorf, T., Grohe, K., Najbauer, E.E., Ehrmann, F.E., Radeva, N., Zitzer, N., Linne, U., Linser, R., Heine, A., Klebe, G.	23%	Publiziert

* Diese Autoren haben einen gleichwertigen Beitrag zur Publikation geleistet

....

Jonathan Cramer Gerhard Klebe

Alles Wissen und alle Vermehrung unseres Wissens endet nicht mit einem Schlusspunkt, sondern mit Fragezeichen. – Hermann Hesse

Contents

Abbreviations

(HR)MS	(High Resolution) Mass Spectrometry
Alloc	Allyloxycarbonyl
Boc	*tert*-Butyloxycarbonyl
BOP	(Benzotriazol-1-yloxy)tris(dimethylamino)-phosphonium Hexafluorophosphate
BPTI	Bovine Pancreatic Trypsin Inhibitor
Cbz	Benzyloxycarbonyl
CSD	Cambridge Structural Database
DCM	Dichloromethane
DEAD	Diethyl Azodicarboxylate
DIPEA	*N*,*N*-Diisopropylethylamine
DMF	Dimethyl Formamide
DMSO	Dimethyl Sulfoxide
DMT	Dimethyl Terephthalate
DPPA	Diphenylphosphoryl Azide
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EP	Endothiapepsin
ERC	European Research Council
ESI	Electron Spray Ionization
EtOAc	Ethyl Acetate
FBLD	Fragment Based Lead Discovery
FOM	Figure Of Merit
GOL	Glycerol

HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HILIC	Hydrophilic–Lipophilic Interaction Chromatography
HOBt	Hydroxybenzotriazole
HOMO	Highest Occupied Molecular Orbital
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
IMPI	Insect Metalloprotease Inhibitor
ITC	Isothermal Titration Calorimetry
LDA	Lithium Diisopropylamide
LUMO	Lowest Unoccupied Molecular Orbital
MAD	Multi-Wavelength Anomalous Diffraction
MD	Molecular Dynamics Simulation
MPD	2-Methyl-2,4-pentanediol
NICS	Nucleus Independent Chemical Shift
NMR	Nuclear Magnetic Resonance Spectroscopy
NOESY	Nucelar Overhauser Effect Spectroscopy
PAINS	Pan Assay Interference Scaffolds
PDB	Protein Databank
PEG	Polyethylene Glycol
PIPES	1,4-Piperazinediethanesulfonic acid
PyBOP	Benzotriazol-1-yl-oxytripyrrolidinophosphonium Hexafluorophosphate
QM	Quantum Mechanics
qNMR	Quantitative Nuclear Magnetic Resonance Spectroscopy
RMSD	Root Mean Square Deviation
RU	Response Unit
SASA	Solvent Accessible Surface Area
SBDD	Structure Based Drug Design

SPE	Solid Phase Extraction
SPR	Surface Plasmon Resonance Spectroscopy
STD	Saturation Transfer Difference
TFA	Trifluoroacetic Acid
THF	Tetrahydrofuran
TLN	Thermolysin

Abstract

In the pre-clinical development stages of most drug design campaigns, the equilibrium binding affinity of a prospective lead candidate, in the form of an IC_{50}, K_d or $\Delta G°$ value, is the most commonly employed benchmark parameter for its effectiveness as a putative drug. Hydrogen bonding, van der Waals and electrostatic interactions, as well as hydrophobic effects are among the most prominent factors that contribute to binding. In structure based design approaches, these interactions can routinely be linked to a structural motif of a drug molecule, which can greatly assist in the construction of compounds with a desired set of properties. Equilibrium binding affinity can also be expressed in terms of kinetics, were the steady-state constant K_d is defined as the ratio of the rate constants of dissociation (k_d) and association (k_a). The thermodynamic expression $\Delta G°$ can be subdivided into an enthalpic ($\Delta H°$) and an entropic ($-T\Delta S°$) term. In either case, the molecular mechanisms that define the kinetics of binding or the compensation of enthalpic and entropic contributions are not fully understood. The goal of this dissertation is the in-depth investigation of the molecular processes that drive protein–ligand interactions. A special focus is set on the partitioning of thermodynamic and kinetic parameters into their respective microscopic elements. For this, the metalloprotease thermolysin (TLN) is used as a model system. This protein is well characterized and represents a robust system with excellent crystallographic properties and a thoroughly documented inhibitor class.

The first publication (Chapter 2) presents an improved strategy for the synthesis and purification of phosphonamidate peptides that are known as potent inhibitors of TLN. Due to the inherent instability of the phosphorous–nitrogen bond, the introduction of polar functional groups into the inhibitor scaffold is quite challenging. Here, a synthetic strategy is presented that minimizes the amount of hydrolysis during peptide coupling, deprotection and purification through the use of an allyl-based protection system and a solid-phase extraction (SPE) protocol for the final purification step. This allows the synthesis of highly pure TLN inhibitors incorporating a variety of functional groups for use in biophysical experiments.

In the second publication (Chapter 3), a strategy for the design of inhibitors is highlighted, which relies on the targeted design of water networks that are formed around a protein–ligand complex. Based on information from a previous study, the shape of a hydrophobic portion of a TLN ligand is altered in a way that allows a beneficial stabilization of water molecules in the

first solvation layer of the complex. Supported by molecular dynamics simulations, a series of diastereomeric inhibitors is synthesized and the binding process is characterized by X-ray crystallography, isothermal titration calorimetry (ITC) and surface plasmon resonance spectroscopy (SPR). The optimization of the hydrophobic P_2' moiety results in a 50-fold affinity enhancement compared to the original methyl substituted ligand. This improvement is mainly driven by a favorable enthalpic term that originates from the stabilization of water polygons in the solvation shell.

In the follow-up study in Chapter 4, the binding signature of a series of inhibitors that place a charged and polar moiety in the solvent exposed S_2' pocket of TLN is investigated. Here, a partially hydrated ammonium group is gradually retracted deeper into the hydrophobic protein environment. From the crystal structures it is evident that the polar ligands do not recruit an increased amount of water molecules into their solvation layer when compared to related analogues that feature a purely aliphatic residue at the solvent interface. The penalty for the partial desolvation of the charged functional group, in combination with the lack of a strongly ordered water network, results in a severe affinity decrease that is driven by an unfavorable enthalpic term.

The deep, hydrophobic S_1' pocket of TLN determines the substrate specificity of the protease and is commonly addressed by high affinity inhibitors. Experimental evidence from previous studies suggests, however, that this apolar crevice is only poorly solvated in the absence of an interaction partner. With the study in Chapter 5, an attempt for the experimental analysis of the hydration state of the S_1' pocket is presented. For this, a special inhibitor is designed that transforms the protein pocket into a cavity, while simultaneously providing enough empty space for the accommodation of several water molecules. A detailed analysis of an experimentally phased electron density map reveals that the cavity remains completely unsolvated and thus, vacuous. As an intriguing prospect for the exploitation of such poorly hydrated protein pockets in drug design, the placement of an *iso*-pentyl moiety in the ligand's P_1' position results in a dramatic, enthalpically driven gain in affinity by a factor of 41 000.

With a detailed structural analysis of a series of chemically diverse TLN inhibitors, the kinetics of the protein–ligand binding process are investigated in Chapter 6. From the SPR derived kinetic information, it becomes apparent that the nature of the functional group in the P_2' position of a thermolysin inhibitor has a significant impact on its dissociation kinetics. This property can be linked to the interaction between the respective functionality of a ligand and Asn112, a residue that lines the active site of the protease and is commonly believed to align a substrate for proteolytic cleavage. This residue undergoes a significant conformational change when the protein transitions from its closed state to its open form, from which a ligand is released. Interference with this retrograde induced-fit mechanism through strong hydrogen-bonding interactions to an inhibitor results in a pronounced deceleration of the dissociation

process. The case of the known inhibitor ZF^{P}LA demonstrates that a further restriction of the rotation of Asn112 by a steric barrier in the P_1 position of a ligand, can reduce the rate constant of dissociation by a factor of 74 000.

Fragment-based lead discovery has become a popular method for the generation of prospective drug molecules. The weak affinity of fragments and the necessity for high concentrations, however, can result in false-positive signals from the initial binding assays that routinely plague fragment-based screening. The pursuit of such a "red herring" can lead to a significant loss of time and resources. In Chapter 7, a molecule that emerged as one of the most potent binders from an elaborate fragment screen against the aspartic protease endothiapepsin is identified as a false-positive. Detailed crystallographic, HPLC and MS experiments reveal that the affinity detected in multiple assays can in fact be attributed to another compound. This entity is formed from the initially employed molecule in a reaction cascade that results in a major rearrangement of its heterocyclic core structure. Supported by quantum chemical calculations and NMR experiments, a mechanism for the formation of the elusive compound is proposed and its binding mode analyzed by X-ray crystallography.

Zusammenfassung

In der präklinischen Phase einer Wirkstoffentwicklung wird häufig die Affinität einer Verbindung im thermodynamischen Gleichgewicht in Form eines IC_{50}, K_d oder $\Delta G°$ Wertes als Referenzparameter für ihre Effektivität als möglicher Wirkstoffkandidat verwendet. Einige der Faktoren, die auf molekularer Ebene zur Affinität einer Verbindung beitragen sind Wasserstoffbrückenbindungen, van der Waals Interaktionen, elektrostatische Wechselwirkungen, sowie hydrophobe Effekte. Mit Hilfe Struktur-basierter Methoden können diese Wechselwirkungen häufig den strukturellen Motiven eines Wirkstoffkandidaten zugeordnet werden. Dadurch kann die gezielte Konstruktion von Molekülen mit gewünschten Eigenschaften ermöglicht werden. Bindungsaffinität kann mithilfe des kinetischen Terms K_d als Quotient der Geschwindigkeitskonstanten von Dissoziation (k_d) und Assoziation (k_a) ausgedrückt werden. Der thermodynamische Ausdruck $\Delta G°$ lässt sich in einen enthalpischen ($\Delta H°$) und einen entropischen Beitrag ($-T\Delta S°$) aufteilen. In beiden Fällen sind die molekularen Mechanismen, die die Bindungskinetik definieren oder zur Kompensation von Enthalpie und Entropie beitragen, nur unvollständig verstanden. Das Ziel der vorliegenden Arbeit ist eine detaillierte Untersuchung der molekularen Prozesse, die die Interaktion von Proteinen und ihren Liganden ausmachen. Ein besonderes Augenmerk liegt dabei auf der Aufspaltung thermodynamischer und kinetischer Größen in ihre jeweiligen mikroskopischen Elemente. Hierfür wird die Metalloprotease Thermolysin (TLN) als Modellsystem verwendet. Dieses Protein ist gut charakterisiert und zeichnet sich daher als ein robustes Testsystem mit exzellenten kristallografischen Eigenschaften und einer bekannten Klasse von Inhibitoren aus.

In der ersten Publikation (Kapitel 2) wird eine verbesserte Strategie für die Synthese und Aufreinigung von peptidischen Phosphonamidaten vorgestellt, die als potente Inhibitoren von TLN bekannt sind. Die inhärente Labilität der Phosphor–Stickstoff Bindung dieser Substanzklasse erschwert die Einführung polarer funktioneller Gruppen in das Inhibitor-Grundgerüst. Mit Hilfe einer neuen synthetischen Methode kann die Hydrolyse der Verbindungen während der Peptidkupplung, Entschützung und Aufreinigung durch die Verwendung einer Allyl-basierten Schutzgruppenstrategie und einer Festphasenextraktionsmethode auf ein Minimum reduziert werden. Dadurch wird die Darstellung von TLN-Inhibitoren mit einer Vielzahl funktioneller Gruppen in hoher Reinheit ermöglicht.

Die zweite Publikation (Kapitel 3) befasst sich mit einer Methode für den Entwurf neuer Inhibitoren, die auf dem gezielten Design des einen Protein–Ligand Komplex umhüllenden Wassernetzwerks basiert. Basierend auf den Ergebnissen einer vorangegangenen Studie wird die Form eines hydrophoben Teils des TLN-Liganden solcherart verändert, dass eine begünstigte Stabilisierung von Wassermolekülen in der Hydrathülle des Komplexes ermöglicht wird. Unterstützt durch Molekulardynamiksimulationen wird eine Serie diastereomerer Inhibitoren synthetisiert und deren Bindungseigenschaften mittels Röntgenkristallstrukturanalyse, isothermaler Titrationskalorimetrie (ITC) und Oberflächenplasmonresonanzspektroskopie (SPR) untersucht. Die Optimierung der apolaren P_2'-Gruppe des Inhibitors resultiert in einer 50-fachen Verbesserung der Affinität im Vergleich mit dem ursprünglichen Methyl-substituierten Liganden. Dieser Gewinn wird hauptsächlich durch einen vorteilhaften enthalpischen Term bedingt, der aus einer Stabilisierung polygonaler Wasserstrukturen in der ersten Hydratationsschicht entstammt.

In der Folgestudie in Kapitel 4 wird die Binding einer Serie von Inhibitoren untersucht, die eine polare und geladene Gruppe in die Lösungsmittel-exponierte S_2' Tasche von TLN platzieren. Eine terminale Ammonium-Funktionalität wird hierbei kontinuierlich tiefer in die hydrophobe Umgebung des Proteins gezogen. Die Untersuchung der Kristallstrukturen zeigt, dass die polaren Liganden, im Vergleich mit unpolaren Analoga, keine verstärkte Nahordnung in der umgebenden Wasserstruktur bewirken. Der Beitrag für die partielle Desolvatation der geladenen Gruppe in Kombination mit der Abwesenheit eines starken Wassernetzwerks hat einen empfindlichen Verlust an Bindungsaffinität, hauptsächlich bedingt durch einen ungünstigen enthalpischen Term, zufolge.

Die tiefe, hydrophobe S_1' Tasche von TLN bedingt die Substratspezifität der Protease und wird häufig von potenten Inhibitoren adressiert. Vorangegangene Experimente legen nahe, dass diese Bindetasche in Abwesenheit eines Interaktionspartners jedoch nur unvollständig hydratisiert ist. Die Studie in Kapitel 5 stellt eine experimentelle Untersuchung des Solvatationszustandes der S_1' Tasche vor. Hierfür wird ein spezieller Inhibitor entwickelt, der die Proteintasche abdeckt. Die so entstandene Kavität bietet weiterhin genug Platz um die Bindung mehrerer Wassermoleküle zu ermöglichen. Die Analyse einer experimentell phasierten Elektronendichtekarte zeigt jedoch, dass die Kavität nicht solvatisiert, und somit vollständig leer ist. Vielversprechend für die Ausnutzung solch unvollständig hydratisierter Taschen für die Entwicklung neuer Arzneistoffe ist die Beobachtung, dass die Platzierung einer *iso*-Pentyl Gruppe in der P_1' Position des Liganden eine dramatische, enthalpiegetriebene Erhöhung der Affinität um den Faktor 41.000 zur Folge hat.

Eine detaillierte Analyse der Kristallstrukturen einer Serie von chemisch unterschiedlichen TLN-Inhibitoren ermöglicht die Untersuchung der Kinetik des Protein–Ligand Bindungsprozesses in Kapitel 6. Anhand der kinetischen Daten aus SPR Experimenten wird

ersichtlich, dass die Art der funktionellen Gruppe in der P_2'-Position des Liganden einen erheblichen Einfluss auf die Dissoziationskinetik aufweist. Diese Eigenschaft kann auf die Interaktion der jeweiligen Funktionalität des Inhibitors mit Asn112, einer Aminosäure, die dafür bekannt ist, ein Substrat für die Peptidspaltung in der Bindetasche auszurichten, zurückgeführt werden. Die Seitenkette dieser Gruppe erfährt eine signifikante Konformationsänderung wenn das Protein aus seiner geschlossenen Form in die offene Konformation übergeht, aus der ein Ligand dissoziieren kann. Eine Beeinflussung dieses Mechanismus durch starke Wasserstoffbrückenwechselwirkungen zu einem Inhibitor führen zu einer Verlangsamung des Dissoziationsprozesses. Aus dem Fall des bekannten Inhibitors $ZF^{p}LA$ wird ersichtlich, dass eine zusätzliche Behinderung der Bewegung von Asn112 durch eine sterische Barriere in der P_1-Position des Liganden die Geschwindigkeitskonstante der Dissoziation um einen Faktor von 74.000 verringern kann.

Die Fragment-basierte Suche nach neuen Leitstrukturen hat sich als eine potente Strategie für die Entwicklung neuer Wirkstoffe erwiesen. Die niedrige Affinität der Fragmente und die in den initialen Assays häufig verwendeten hohen Konzentrationen begünstigen jedoch das Auftreten falsch positiver Ergebnisse. Die Verfolgung einer solchen „falschen Fährte" kann mit einem erheblichen Verlust von Zeit und Ressourcen verbunden sein. In Kapitel 7 wird die Affinität eines Moleküls, das in einem Fragment-basierten Screening gegen die Aspartylprotease Endothiapepsin als hochpotenter Binder aufgefallen war, als falsch-positives Ergebnis identifiziert. Umfangreiche Kristallografie-, HPLC- und MS-Experimente zeigen, dass die Bindungseigenschaften, die in mehreren unabhängigen Methoden detektiert wurden, tatsächlich einer anderen Verbindung zugeordnet werden können. Dieses Molekül wird aus der initial eingesetzten Substanz in einer Reaktionskaskade gebildet, die mit einer massiven Umlagerung des heterozyklischen Grundgerüsts einhergeht. Unterstützt durch quantenmechanische Berechnungen und NMR-Experimente wird ein Mechanismus für die Bildung der Verbindung postuliert und deren Bindungsmodus mittels Röntgenkristallografie aufgeklärt.

Chapter 1

General Introduction and Thesis Outline

1.1 The Study of Molecular Recognition in Rational Drug Design

1.1.1 The Paradox of Modern Drug Development

Modern pre-clinical drug development programs are often based on the screening of large compound libraries, comprising thousands or even millions of entries. The binding affinity of compounds in such libraries to a target protein of interest can routinely be determined in highly automated biochemical or biophysical experiments. In combination with a method that provides information about the binding mode of a ligand, these affinity parameters are then used to guide the optimization process. Here, techniques that give three dimensional structural information, such as X-ray crystallography, are particularly helpful to identify possible interaction vectors. While the determination of a protein structure by crystallography or NMR techniques historically represented a significant experimental challenge, modern methods have dramatically simplified this process. In many cases, improvements like powerful recombinant expression systems, as well as automated pipetting and plate monitoring devices for crystallography allow the determination of a protein structure with a manageable experimental effort. Owed to the broad availability of modern synchrotron radiation facilities for academic and industrial research, the quality of the collected crystallographic data is higher than ever. In addition, the advent of computer-based methods in drug design yielded algorithms that are able to predict the binding pose of a ligand or even its binding affinity and guide the development of a drug candidate.

Yet, despite the rapid progress of experimental and computational tools in rational drug design, the number of new molecular entities that enter the market is in steady decline. In many cases, lead compounds that perform well in pre-clinical settings fail to achieve a relevant pharmacological effect when applied to a patient population in clinical trials. Molecules that prove useful in *in vitro* assays are abandoned, because of toxicity or unfavorable pharmacokinetic properties. Compounds that are scored high by computer algorithms ultimately fail to show activity when they are tested in an experimental setting.

From these drawbacks it is evident that our understanding of the molecular processes that drive the interaction of proteins and ligands in biological recognition processes is still incomplete. Additional basic research is required to elucidate the molecular mechanisms that contribute to the effectiveness of successful drugs and identify reasons for the failure of promising lead compounds in clinical settings.

1.1.2 Dissection of Common Affinity Metrics

Most optimization programs in medicinal chemistry rely on a measure of equilibrium binding affinity for the evaluation of a prospective lead compound. Commonly employed equilibrium

binding metrics, such as K_d or $\Delta G°$ values, can be subdivided into microscopic elements. The kinetic term K_d is defined as the quotient of the rate constants of dissociation (k_d) and association (k_a). The thermodynamic expression $\Delta G°$ can be divided into an enthalpic ($\Delta H°$) and an entropic ($-T\Delta S°$) contribution. These relationships are of particular interest for drug design, since some of the parameters have been linked to beneficial properties of drug molecules. A slow dissociation rate, for example, is believed to improve the efficacy of a compound in an *in vivo* setting and inhibitors with a predominantly enthalpic contribution to binding affinity have may have the potential to show a higher selectivity against off-target effects, as well as a higher tolerance to resistance mechanisms[1,2].

However, while medicinal chemistry programs are routinely able to improve the equilibrium binding affinity of a lead molecule, a prediction of the distribution of affinity values into their constituents has been proven to be difficult. It is apparent that the molecular determinants that define biological recognition are still poorly understood. To further our understanding of the underlying processes, it is necessary to investigate protein–ligand interaction with thoroughly characterized test systems under well-defined experimental conditions. Only when individual parameters can be carefully singled out, a reliable evaluation of the intricate processes, which are often superimposed by a plethora of other effects, is possible within the complex setting of biological system.

1.2 Thermodynamics in Drug Development

1.2.1 The Binding Thermodynamics of Drug Molecules

The determination of thermodynamic values adds another layer of complexity to the data evaluation in the context of a drug design effort. Apart from the overall affinity of a prospective lead compound, thermodynamic data give information about the molecular processes that drive the binding of a respective ligand [2]. Electrostatic and dipolar interactions, hydrogen bonding and van der Waals interactions are factors that contribute to the enthalpic term, whereas the entropic term is influenced by the gain or loss of rotational or translational degrees of freedom and the displacement of conformationally restricted water molecules into the bulk solvent. However, it is important to note that the data derived from experiments such as isothermal titration calorimetry (ITC) comprise contributions of the whole binding event, starting from the separately solvated interaction partners until the system is at equilibrium. Any process that gives rise to a heat signal during the course of that reaction contributes to the resulting shape of the binding isotherm. These include effects like desolvation, conformational transitions, changes in protonation state or the residual water solvation pattern surrounding any of the participating entities. In addition, a frequently observed compensation of enthalpic and entropic contributions, resulting in a similar free energy of binding, further complicates

data analysis [3,4]. A meaningful correlation of distinct structural features of a ligand to thermodynamic properties requires a careful analysis of a congeneric series of similar compounds in combination with high quality crystallographic information [5].

1.2.2 Solvent Effects on the Thermodynamics of Molecular Recognition

The solvation shell of the interaction partners in a biological system plays an active role in any binding process. Before two species can directly interact with one another, water molecules that form the hydration layer around either binding partner have to be expelled from the solute–solvent complex. The displacement of water from a hydrophobic surface patch into the bulk water phase is commonly associated with an entropic advantage (see 1.2.3). In contrast to that, hydrophilic groups engage in strong polar interactions to surrounding water molecules. If solute–solvent hydrogen bonds have to be broken to allow an interaction of a ligand with a target protein, the resulting enthalpic penalty has to be compensated by equally favorable contacts between the two solutes. The desolvation penalty for any hydrogen bond that is not saturated in the ligand's binding pose represents a severe loss in affinity. These findings underline why solvation effects have to be considered in the development of a prospective drug molecule. Especially if polar functional groups are introduced into a ligand scaffold to improve its solubility or pharmacokinetic properties without engaging in meaningful interactions to a protein, an increased desolvation penalty can easily result in a significant loss of affinity.

1.2.3 Hydrophobic Effects in Molecular Recognition

A dominant role in molecular recognition in water has been attributed to the so called "hydrophobic effect". This expression describes the entropic benefit that arises from a displacement of orientationally constrained water molecules at a hydrophobic interface into the bulk solvent [6,7]. This effect is not only cited as the reason for the phase separation of an oil–water system, it has also been implicated as the main driving force behind protein folding, aggregation, and protein–ligand interaction [8]. However, detailed analyses of the thermodynamics of hydrophobic hydration have also shown that this simplified model does not hold true in every case. In multiple instances, the displacement of water molecules from an apolar surface has been reported to give rise to an enthalpically favorable change in free energy [9–11]. These "non-classical hydrophobic effects" have been attributed to the displacement of highly mobile, disordered water molecules in hydrophobic environments, whose entropy content does not differ significantly from bulk. The energetic benefit upon their release is consequently driven by enthalpically beneficial solvent–solvent interactions, as well as newly formed van der Waals contacts between the hydrophobic solutes. The hypothesis that water

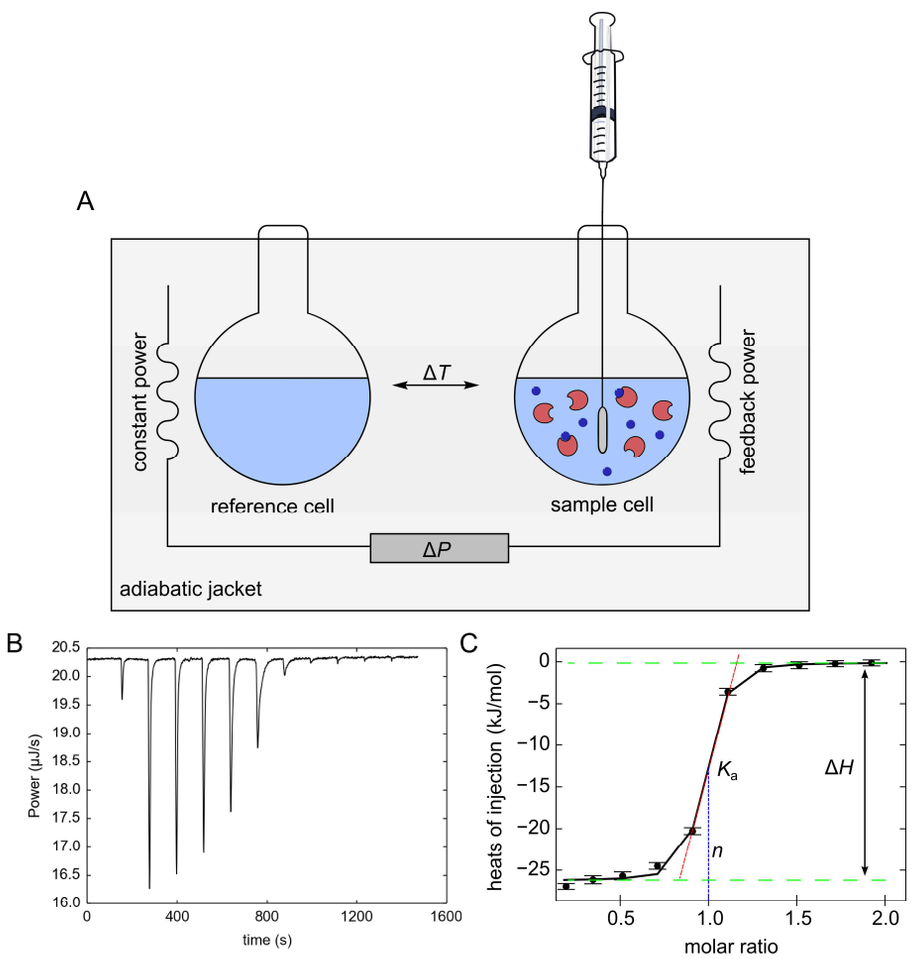


Figure 1.1. Typical setup of an ITC instrument. A) The ITC apparatus determines the heat content that is generated by a binding event in the sample cell in comparison with a reference cell. The power that is necessary to maintain the isothermal condition over time is recorded. B) Thermogram recorded during the course of a titration. C) Binding isotherm extracted from a typical ITC experiment. The enthalpy of binding is represented by the distance between the plateaus of the sigmoidal curve. The association constant K_a can be derived from slope at the inflection point. The position of the inflection point on the molar ratio axis gives information about the site number *n*.

molecules that are located at hydrophobic interfaces generally assume a more ordered, entropically unfavorable state, could not be proven experimentally[12,13]. This highlights the fact that the energetic details of solvation are not uniform, but instead are very much dependent on the local features, especially the shape, of a given binding site and the resulting structure of water molecules in its solvation layer [14–16]. Therefore, the common usage of the term "hydrophobic effect" is often misleading. Instead, it has to be considered that favorable hydrophobic effects can originate for diverse thermodynamic reasons to a very different extent.

1.2.4 Determination of Binding Thermodynamics by ITC

In experiments with biological systems, thermodynamic parameters are most commonly determined by ITC (Figure 1.1). This method allows the simultaneous assessment of K_a, $\Delta G°$, $\Delta H°$ and $-T\Delta S°$, as well as the "*site number*" n, a descriptor of the interaction mechanism or the stoichiometry of the reaction [17]. An ITC instrument records the differential power that is necessary to maintain a constant temperature between a reference cell and a sample cell, after a defined amount of an interaction partner has been titrated into the sample cell, which contains an analyte of interest. In this way, the heat signal from the interaction between the two species is determined, which allows the calculation of the parameters n, K_a and $\Delta H°$ during the course of a complete titration. The terms $\Delta G°$ and $-T\Delta S°$ are in turn derived from the experimentally determined values.

1.3 Binding Kinetics in Drug Development

1.3.1 Drug–Target Residence Time

Most biochemical and cellular assays that are used to investigate the interaction of a prospective drug with its target protein are based on the determination of the binding affinity under thermodynamic equilibrium conditions. Most commonly, *in vitro* experiments for the determination of IC_{50}, K_d or $\Delta G°$ values operate under "closed system" conditions with fixed concentrations of the interaction partners. In stark contrast to that, *in vivo* pharmacology is characterized by non-equilibrium conditions. In this "open system", the concentration of an applied drug is constantly affected by pharmacokinetic processes like absorption, metabolism and excretion. For this reason, the predictive properties of equilibrium binding metrics for *in vivo* efficacy have been questioned [18,19]. The fixation on these, arguably inappropriate, descriptors in preclinical development has been deemed a probable cause for the high rate of attrition in the later stages of clinical trials. As an alternative predictor for the effectiveness of a prospective drug, the lifetime of the protein–ligand complex has been suggested [18,20–24]. The rationalization behind this hypothesis is the assumption that the pharmacological effect of a drug can only persist as long as the molecule is bound to its target. Only when the ligand dissociates from its binding site, the addressed protein is able to execute its pathophysiological function.

The lifetime of a protein–ligand complex is determined by the rate of association and the rate of dissociation. Since the rate constant of association is subjected to several physicochemical and pharmacological limitations in an *in vivo* setting, complex lifetime can typically be described with the rate constant of dissociation (k_d) or related parameters like the half life of dissociation ($t_{½}^{diss} = \ln 2\ k_d^{-1}$) or, more commonly, residence time ($\tau = k_d^{-1}$). These parameters can easily be determined by SPR or another suitable method in a preclinical setting.

While this approach is not without debate, multiple studies have been reported that support a correlation between drug–target residence time and *in vivo* efficacy [25–27].

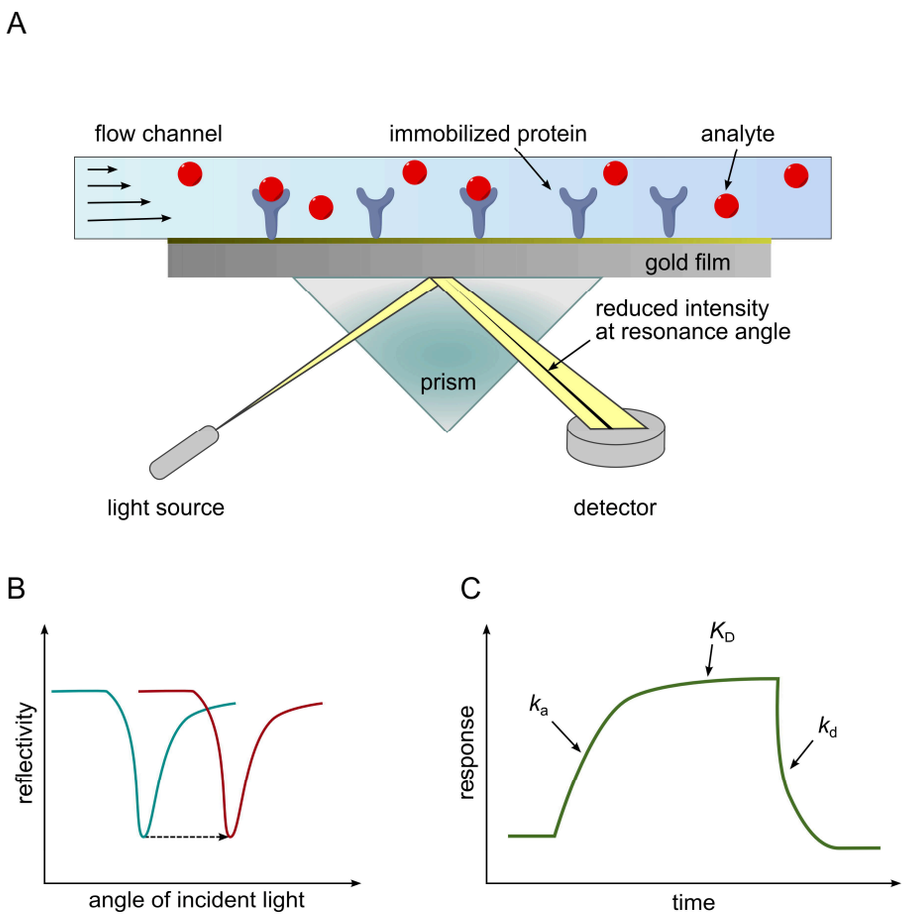


Figure 1.2. Typical setup of an SPR instrument. A) A solution of an analyte is passed over the gold surface of a sensor chip. The binding event to a protein, which is immobilized on the surface, changes the local refractive index. The resulting variation in the resonance angle of the surface plasmon is detected by an optical setup. B) A shift of the angle of total reflection indicates a binding event. C) The plot of SPR response against time can be fit to a kinetic model, from which the rate constants can be calculated.

The molecular determinants that define the binding affinity of a molecule have been studied for decades. By optimizing the relevant interactions of a ligand, it is typically possible to achieve a significant gain in affinity during the course of a lead optimization program. In contrast to that, a profound understanding of how these parameters split up into association and dissociation kinetics is still lacking. While the focus on long residence time inhibitors represents a promising strategy for drug design, a targeted optimization has proven to be difficult, because the mechanisms that contribute to the lifetime of a protein–ligand complex are not fully understood in most cases.

1.3.2 Determination of Binding Kientics by SPR

In theory, the kinetics of a protein–ligand interaction can be determined by any method that is able to monitor and differentiate the concentrations of the bound and unbound state of a protein in a time resolved manner. However, the use of a technique that relies on the optical phenomenon of surface plasmon resonance (SPR, Figure 1.2) has constantly gained in popularity for this application [28–30]. This method is able to detect changes in the refractive index that occur when molecules adsorb to or desorb from a solid–liquid interface. When a protein of interest is immobilized on that surface, the association and dissociation of a binding partner can be monitored in very high temporal resolution. The shape of the resulting sensogram can be fit to a kinetic model of the interaction, which provides information about the rate constants of association (k_a) and dissociation (k_d), as well as the steady-state dissociation constant K_d.

1.4 The Metalloprotease Thermolysin as a Model System for Medicinal Chemistry

Thermolysin (TLN) is a zinc-dependent metalloprotease from *B. thermoproteolyticus* that belongs to the M4 family of proteases. The protein consists of 316 amino acid residues and commonly binds four calcium ions that are responsible for its profound thermostability [31]. The tertiary structure of the enzyme can be divided into a mostly helical domain at its *C*-terminus and an *N*-terminal domain that features a distinct β-sheet motif (Figure 1.3). The active site is located on a central α-helix in between the two domains. The characteristic HEXXH-motif, which commonly binds the catalytic zink ion within the M4 family and related metalloproteases, is formed by His142, Glu 143, and His146 [32].

The substrate specificity of TLN is mainly determined by its well defined hydrophobic S_1' cavity. Here, apolar residues of moderate and large size are accommodated, which results in a preference for substrates with Val, Leu, or Phe residues in the P_1' position. The non-primed S_1 and S_2 sites, as well as the less defined S_2' pocket, are mostly apolar and solvent-exposed. None of these binding sites contribute significantly to the affinity against a substrate or inhibitor [33].

The first crystal structure of TLN has been reported in 1972 [34]. From this point on, the enzyme has been utilized as a model system for the development of inhibitors against pharmaceutically relevant targets such as angiotensin-converting enzyme or neutral endopeptidase [35–37]. Its chemical robustness, excellent crystallographic properties and ready access in large quantities also disclose TLN as an ideal test system to study the details of protein–ligand interactions.

1.5 Phosphonamidate Peptides as Inhibitors of Thermolysin

1.5.1 Phosphorous Containing Inhibitors of Thermolysin

The phosphorous containing natural product phosphoramidon (**1**) from *S. tanashiensis* has been described as a potent inhibitor of TLN and related proteases (Figure 1.4) [38]. This compound features a central tetraedric phosphoramidate moiety that is able to coordinate to the catalytic zink ion and mimic the transition state of peptide cleavage. Based on the same mechanism of action, the related group of phosphonamidate peptides (**2**) have been introduced. These compounds have been proven to be potent inhibitors of proteases such as TLN [39], Carboxypeptidase A [40], angiotensin-converting enzyme [41], or endothelin-converting enzyme [42], and have been successfully used as valuable tools for the investigation of protein–ligand interactions.

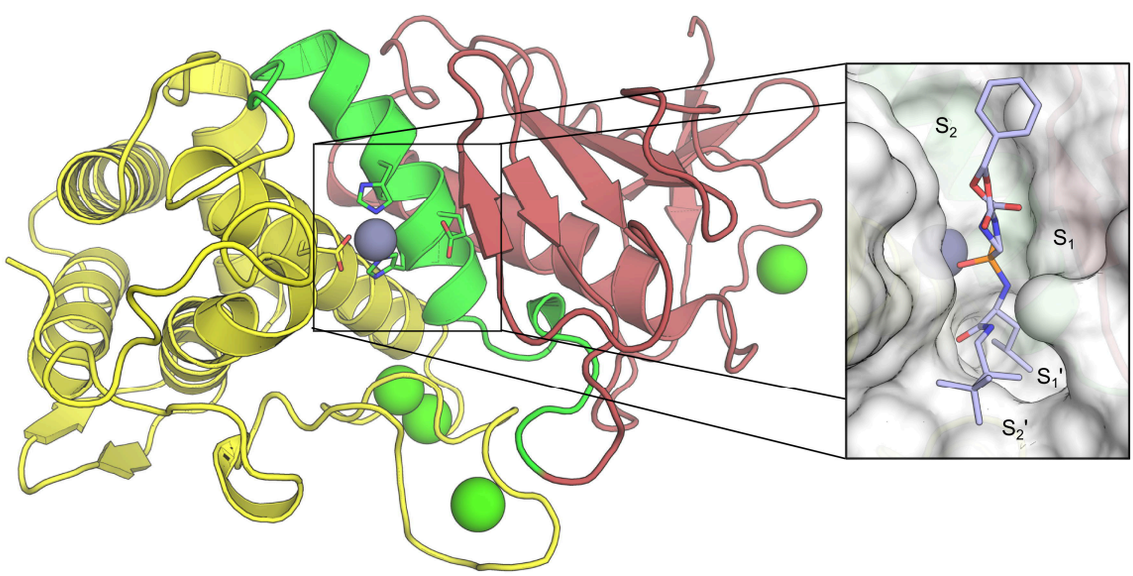


Figure 1.3. Cartoon representation of TLN in the closed state (PDB 5MNR). The central helix (green) is connected to the helical *C*-terminal domain (yellow) and the *N*-terminal domain (red) that is characterized by a distinct β-sheet motif. The catalytic zink ion (blue sphere) is coordinated by His142, His146 and Glu166. Glu143 is retracted into the S_1 pocket in the closed state of the enzyme. Four calcium ions (green spheres) in the peripheral loop regions contribute to the remarkable stability of TLN. The small image section shows an inhibitor bound to the active site of TLN (PDB 5JS3). The interaction is driven by the accommodation of a hydrophobic residue to the S_1' specificity pocket. An interaction with the shallow and solvent exposed S_1, S_2, and S_2' pockets has a weaker contribution to binding affinity.

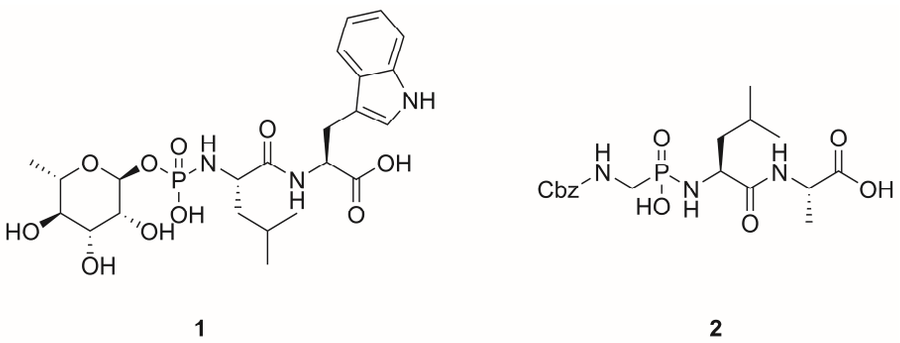
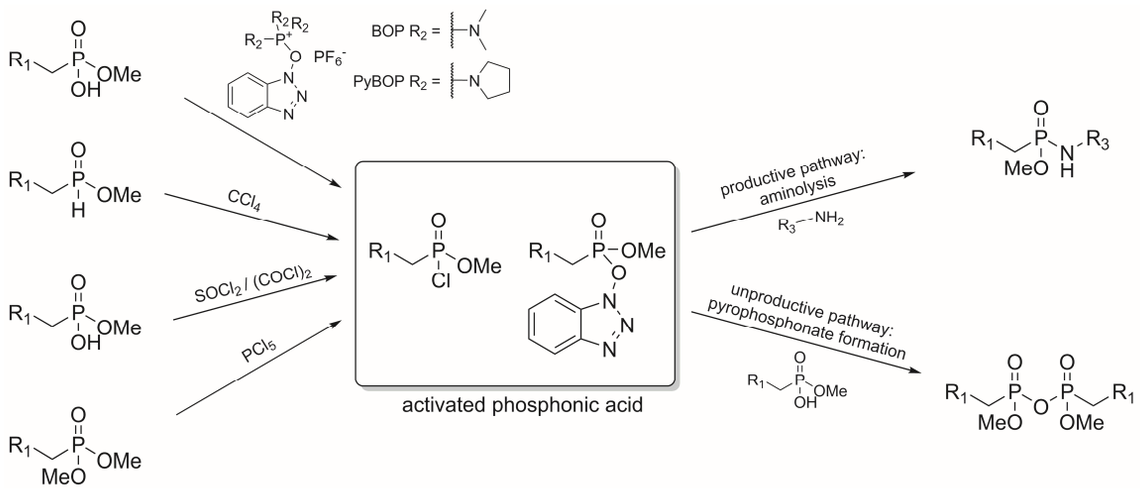


Figure 1.4. Phosphorous containing inhibitors of TLN.

1.5.2 **Physicochemical Properties of Phosphonamidates**

Despite the obvious relation to carboxamides, the physicochemical properties of phosphonamidates differ significantly from those of their carbon-analogs. The defining feature of carboxamides is the partial delocalization of the nitrogen lone-pair. Quantum-chemical calulations revealed that the carboximidic acid tautomer of acetamide contributes with 28% to the resonance hybrid of the molecule [43]. This property of carboxamides results in a rotational barrier of 60–80 kJ mol^{-1} around the carbon–nitrogen bond. In addition, a pK_a value of 15–24 is determined for the N-H protons of primary and secondary carboxamides [44,45]. With a pK_a value of –1, these compounds are preferably protonated at the carbonyl oxygen. In contrast to that, entirely different properties have been reported for phosphonamidates. A pK_a value of 5 is commonly determined for a free phosphonamidic acid [46]. This value is about three orders of magnitude higher than the expected K_a of a typical phosphonic acid. Supported by crystallographic experiments and NMR data, this observation endorse the assumption that phosphonamidates are preferably protonated at the phosphonamidate nitrogen and commonly exist in a zwitterionic state [46].

This property has major implications for the stability of phosphonamidates, especially in aqueous solution. The charged ammonium group of an *N*-protonated phosphonamidate can act as a leaving group in a nucleophilic substitution reaction. This explains the fact that a half life of several minutes to hours in aqueous acidic solution has been found for some compounds of this class [40,46–48].



Scheme 1.1. Commonly employed strategies for the activation of phosphonic acids in phosphonamidate coupling reactions.

1.5.3 Synthesis of Phosphonamidates

The most commonly employed strategy for the synthesis of phosphonamidates is the activation of a phosphonic acid as a phosphonochloridate. These intermediates can either be prepared from phosphonic acid diesters by treatment with PCl_5 [49] or from a reaction of phosphonic acid monoesters with thionyl or oxalyl chloride [50,51]. Another synthetic access to phosphonochloridates is the Atherton-Todd reaction, an oxidative chlorination of phosphinic acid esters with CCl_4 [52]. The use of modern peptide coupling reagents has also been explored for the formation of the P–N bond in phosphonamidates [49,53–56]. While compounds like carbodiimides or uronium reagents routinely give excellent results for the synthesis of carboxamides, their performance as coupling reagents for phosphonamidates has been notoriously mediocre. The best results have been achieved with the phosphonium reagents BOP and PyBOP (Scheme 1.1).

The reason for the sluggish reaction rates of phosphonamidate coupling is the formation of several inactive species during the complex reaction mechanism. Especially pyrophosphonate intermediates represent a dead end for the infamously slow reaction of phosphonic acid derivatives with amines [50,54]. However, it has been shown that a profound knowledge of this mechanism and an adaption of the experimental procedure to the specific requirements of the reaction does in fact allow the use of most coupling reagents for the formation of phosphonamidates [56].

1.6 Fragment-based Lead Discovery

1.6.1 General Methods

Fueled by impressive reports of success, fragment-based lead discovery (FBLD) methods have massively gained in popularity since their emergence in the 1990s [57,58]. These methods rely on the screening of so called fragments, typically molecular entities with a formula weight of less than 300 g mol^{-1}, for affinity against a target protein. The entries of a fragment library have to be diligently selected, so that a broad spectrum of interaction capabilities, three dimensional shape, and complexity are covered. In addition to that, the library should provide diverse synthetic vectors for later elaboration of a fragment into a lead molecule [59,60]. Following the famous concept of Lipinski's "rule of five" for drug design, a similar "rule of three" has been proposed as a guidance for fragment sized libraries [61]. When these considerations are adequately addressed, a fragment library is able to cover a significantly larger portion of the available chemical space compared to a full sized compound library for high throughput screening.

The methods that are commonly used for affinity screening consist of classical biochemical assays that utilize the detection of fluorescence or radioligand binding, as well as modern

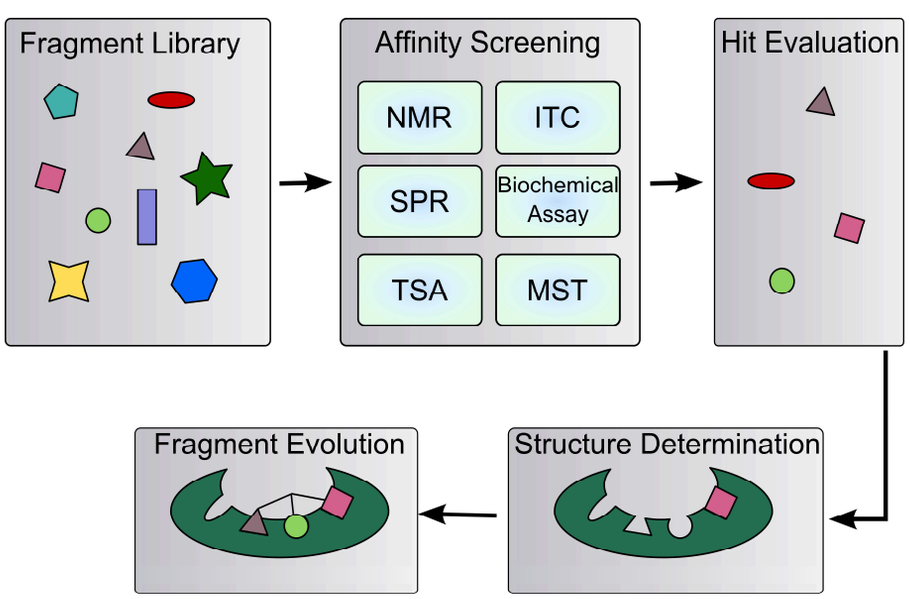


Figure 1.5. Typical Workflow of fragment-based lead discovery. A library of fragment-sized compounds is screened for its affinity against a protein of interest. The binding mode of the resulting hits is determined by X-ray crystallography. The original fragment hit is evolved into a full sized lead compound by a combination with other screening hits (fragment merging, fragment linking) or a structure-based drug design approach (fragment growing).

biophysical techniques, such as thermophoresis, thermal shift assay, NMR, SPR, and ITC. In combination with a method that gives information about the binding mode of a prospective hit, a comprehensive map of the interaction capabilities within the binding site of a target protein can be constructed.

Taking advantage of this information, the fragment-sized screening hits can then be elaborated into full sized lead compounds. This can be achieved by merging or linking several fragments together, or by applying structure-based techniques to grow a molecule to drug-like proportions. In either case, the binding properties of the resulting lead compounds usually exceed the affinity of the initial hits by several orders of magnitude (Figure 1.5).

1.6.2 Evaluating the Performance of Pre-Screening Assays

The initial selection of molecules from a compound library is most commonly based on the results from a biochemical or biophysical assay that gives information about the binding affinity of a fragment. Sometimes, a screening cascade is employed that utilizes an initial, fast and crude method for the primary selection of a library subset, followed by a more elaborate technique [62,63]. The goal of these pre-screening efforts is the selection of a set of compounds that are subsequently analyzed by putatively low-throughput methods like protein X-ray crystallography or three-dimensional NMR techniques, in order to obtain information about their binding mode and viable interaction vectors within the protein binding pocket. Employing the aspartic protease endothiapepsin (EP) from the fungus *C. parasitica*, an elaborate study was set up to investigate differences in the yield of a variety of modern screening techniques that are routinely used for the evaluation of fragment libraries [64,65]. Surprisingly, the overlap between different pre-screening assays proved to be alarmingly low, which undermines the confidence that an exhaustive set of binders can be reliably extracted from a library by employing a single affinity assay, instead of a cascade of several subsequent assay determinations. Alternatively, a high-throughput crystallographic pipeline was set up that proved to be most efficient for the identification of compounds as a starting point for fragment elaboration [66–68].

1.6.3 False Positives and Pan Assay Interference Compounds

Library screening campaigns have been a fundamental part of drug design efforts in pharmaceutical companies for the last decades. However, data interpretation of high throughput screening results has been notoriously plagued by the occurrence of false positive hits from the initial assays [69]. In many cases, the originally detected binder shows no affinity in subsequent experiments or proves to bind through a non-specific mechanism to the protein

of interest. Metal chelation [70], redox activity [71], chemical reactivity [72,73], aggregation [74], compound impurity [75–77], promiscuous binding [70,78,79], as well as an inherent compound fluorescence or other means of interference with the assay detection method [80] have been identified as reasons for the emergence of false positives. Molecules and functionalities that are infamous for their properties to give false positive readouts in commonly employed assays have been dubbed "PAINS", short for "**P**an **A**ssay **In**terference Compound**s**" [81]. The pursuit of false positive hits from a library screen can represent a significant loss of time and resources. Therefore, the prevention or early detection of such events has become a major concern in screening campaigns.

1.7 Thesis Outline

The work presented in this thesis tries to answer fundamental questions about the molecular processes that drive the interaction of proteins and their binding partners in biological recognition. For this research effort, the metalloprotease thermolysin serves as a reliable and well documented model system. The main objective was the dissection of thermodynamic and kinetic affinity descriptors into their respective elements and a correlation of these values to structural observations. For each specific question, a defined, congeneric series of inhibitors was analyzed with an experimental setup that was tailored to the needs of the respective research objective.

The results presented in Chapter 2 serve as the basis for the following experiments. Here, a reliable synthetic method for the preparation of phosphonopeptidic inhibitors for TLN is detailed. The high polarity and instability of deprotected phosphonamidates complicates their preparation, especially during the final deprotection step. Traditional protocols, such as a purification by preparative reversed phase HPLC, can lead to a decomposition of the target compounds in the aqueous mobile phase. This is particularly prevalent when polar and basic functional groups are included in the inhibitor scaffold. The protocol presented in this section describes a methodology that is based on an alloc/allyl ester protection strategy in combination with a solid phase extraction procedure. This workflow gives the deprotected inhibitors in good yields and excellent purity.

The research described in Chapters 3, 4, and 5 is concerned with the influence of water molecules on the interaction of TLN with its inhibitors. As a solvent that is ubiquitously present in biological systems, water takes an active role in biological recognition processes. Using the TLN system as a case study, the experiments conducted here try to elucidate the nature and extent of solvent effects on the thermodynamics and kinetics of binding. In Chapter 3, the influence of hydrophobic hydration around a solvent-exposed portion of an inhibitor on the thermodynamic signature of binding is investigated. Based on a previously conducted

study, a design concept is evaluated that aims to optimize the structure of water networks in the hydration shell of the ligand. In contrast to this, the experiments presented in Chapter 4 study the effects that charged and polar groups can have on the binding thermodynamics when they are placed in a solvent-exposed region at the surface of the protein–ligand complex. Here, the interactions of a charged group to surrounding water molecules seem to be of a more transient nature, so that the penalty for the partial desolvation of the polar group dominates the binding profile. In Chapter 5, it is attempted to experimentally validate the hypothesis that the deep and well defined hydrophobic S_1' pocket of TLN remains unsolvated when no ligand is bound to the protein. Using a tool compound, which was specifically designed to generate an empty cavity in the binding pocket, in combination with a careful analysis of an experimentally phased electron density map, it is concluded that the S_1' pocket in TLN is indeed empty. The dramatic gain of affinity by a factor of 41 000 for compounds that address this void with an *iso*-pentyl moiety can be attributed to a reduction of the empty space in the cavity.

The focus of the study in Chapter 6 laid on the molecular mechanisms that are responsible for long residence time binding of inhibitors of TLN. Here, binding kinetic data of a series of TLN inhibitors is analyzed by SPR and the information gathered from these experiments is correlated to the crystal structures of the complexes. It is concluded that the rate-determining step of ligand dissociation is correlated to the movement of Asn112, a residue that lines the active site of the protease and is believed to orientate a substrate for peptide cleavage. By interfering with this retro-induced fit mechanism, the residence time of an inhibitor can be dramatically prolonged.

Chapter 7 outlines a side project that is not strictly connected to the questions raised in the previous chapters. In this section, the intrinsic reactivity of a compound is investigated that stood out in a fragment library screening for affinity to the aspartic protease endothiapepsin. The molecule that is found within the active site of the protein in the crystal structure is generated from the originally employed fragment in a reaction cascade that involves an extensive rearrangement of its heterocyclic core structure. The newly generated species binds to the protein by pronounced electrostatic interactions that perfectly complement the topology of the binding pocket. This, potentially highly potent, binder presumably accounts for the high affinity that was detected in initial binding assays.

Chapter 2

An Allyl Protection and Improved Purification Strategy Enables the Synthesis of Functionalized Phosphonamidate Peptides

Jonathan Cramer,[1] and Gerhard Klebe[1]

[1]Department of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany

Synthesis **2017**, 49, 1857–1866

2.1 Abstract

For modern biophysical methods such as isothermal titration calorimetry, high purity of the inhibitor of interest is indispensable. Herein, we describe a procedure for the synthesis and purification of functionalized phosphonamidate peptides that is able to generate inhibitors for the metalloprotease thermolysin for use in biophysical experiments. The method utilizes an allyl ester/alloc protection strategy and takes advantage of a fast and effective solid phase extraction (SPE) purification step. Applying this strategy, we were able to synthesize a series of highly polar inhibitors featuring amino- and hydroxy-functionalized side chains in excellent purity.

Keywords

thermolysin, metalloprotease, inhibitor synthesis, phosphon-amidate, solid phase extraction, allyl/alloc protection

2.2 Main Text

The introduction of phosphonamidate moieties in peptidic structures has repeatedly yielded highly potent transition state analogous inhibitors for a broad range of proteases such as thermolysin (TLN) [39], HIV protease [82], carboxypeptidase A [40], angiotensin-converting enzyme [41], leucine aminopeptidase [83], endothelin converting enzyme [42] and human collagenase [84]. In form of phosphonamidate ester conjugates, the phosphonamidate group has also been incorporated in haptens designed for the generation of catalytic antibodies [85] and utilized for the bio-orthogonal modification of proteins [86]. This scope of applications highlights the applicability of phosphonamidates in biochemical and biomedical research. While phosphonamidate esters are sufficiently stable under physiological conditions [86–88], the inherent instability of free phosphonamidates has been deemed as major disadvantage and prevented a broad application of these compounds [89–91].

In our group, P(V)-based peptidomimetic inhibitors of thermolysin have proven to be valuable tool compounds to study the formation of contiguously connected water networks on the surface of the protein-ligand complex and their influence on binding thermodynamics [15,92–95]. In this context, a high purity of the inhibitors used for the measurements is of vital importance. In isothermal titration calorimetry experiments, the enthalpic term is strongly dependent on the applied inhibitor concentration. Inaccuracies in the latter property can easily lead to false data interpretation when studied samples are not sufficiently pure [96,97].

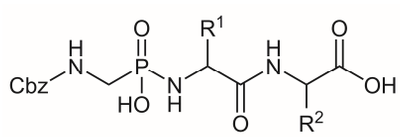
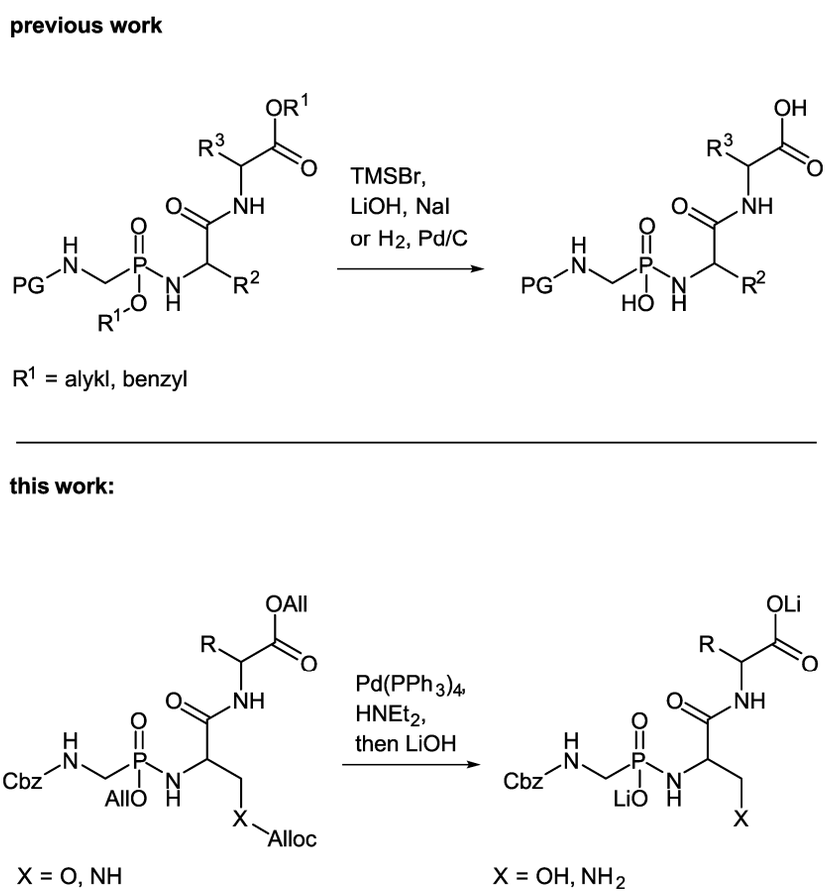


Figure 2.1. Basic scaffold of phosphonamidate type inhibitors of thermolysin.

Herein we report a modified strategy for the synthesis and purification of Cbz-protected phosphonamidates (Figure 2.1), utilizing allyl ester/alloc protection and an efficient solid-phase extraction (SPE) protocol that allows the facile introduction of a variety of functional groups in the P1'- and P2'-position of a peptidomimetic inhibitor. For a prospective biophysical study, we envisaged to introduce polar functional groups to the basic scaffold of thermolysin inhibitors, while simultaneously adhering to the N-terminal Cbz-protecting group, which has to be seen as an integral part of the inhibitor scaffold. The most commonly employed strategy in the synthesis of phosphonamidate peptides utilizes alkyl or benzyl esters as protecting group for the phosphonamidate and C-terminal carboxylic acid functions. The

deprotection is consequently accomplished by hydrolytic cleavage [15], treatment with TMSBr [98], NaI [89] or hydrogenolysis [41], respectively (Scheme 2.1). These strategies combine the advantage of mild reaction conditions with easily removable reagents, thus avoiding the formation of major side products. For the solid-phase synthesis of aryl phosphonamidate peptides by a Staudinger reaction, the necessity of protecting groups can even be eliminated [99]. The purification of the delicate deprotected phosphonamidates is usually achieved by employing ion exchange chromatography or preparative RP-HPLC. However, in various cases, the samples are subsequently used in biochemical assays without any final purification step [83].

Due to the incompatibility of the desired carboxybenzyl phosphonopeptides to hydrogenolytic side chain deprotection and the unsatisfactory range of hydrolytically labile protecting groups for functional groups such as amines, we investigated different strategies as alternative to established protocols. Under consideration of orthogonality and ease of purification after

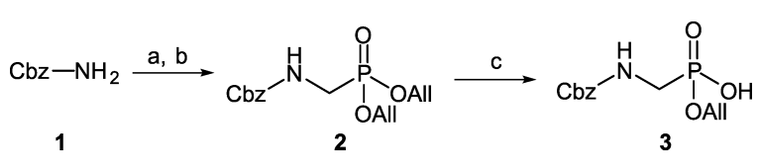


Scheme 2.1. Synthetic strategies for the deprotection of phosphonamidate protease inhibitors.

deprotection, an allyl ester/alloc-strategy was chosen for side-chain and backbone protection. Through usage of diethylamine as nucleophilic scavenger in the palladium-catalyzed deallylation reaction, all reagents and side products, except for the catalyst, are volatile and can be easily removed under reduced pressure. The considerable difference in polarity between the palladium catalyst and the deprotected phosphonopeptides allows for facile separation of the two species by liquid-liquid extraction or reversed phase chromatography.

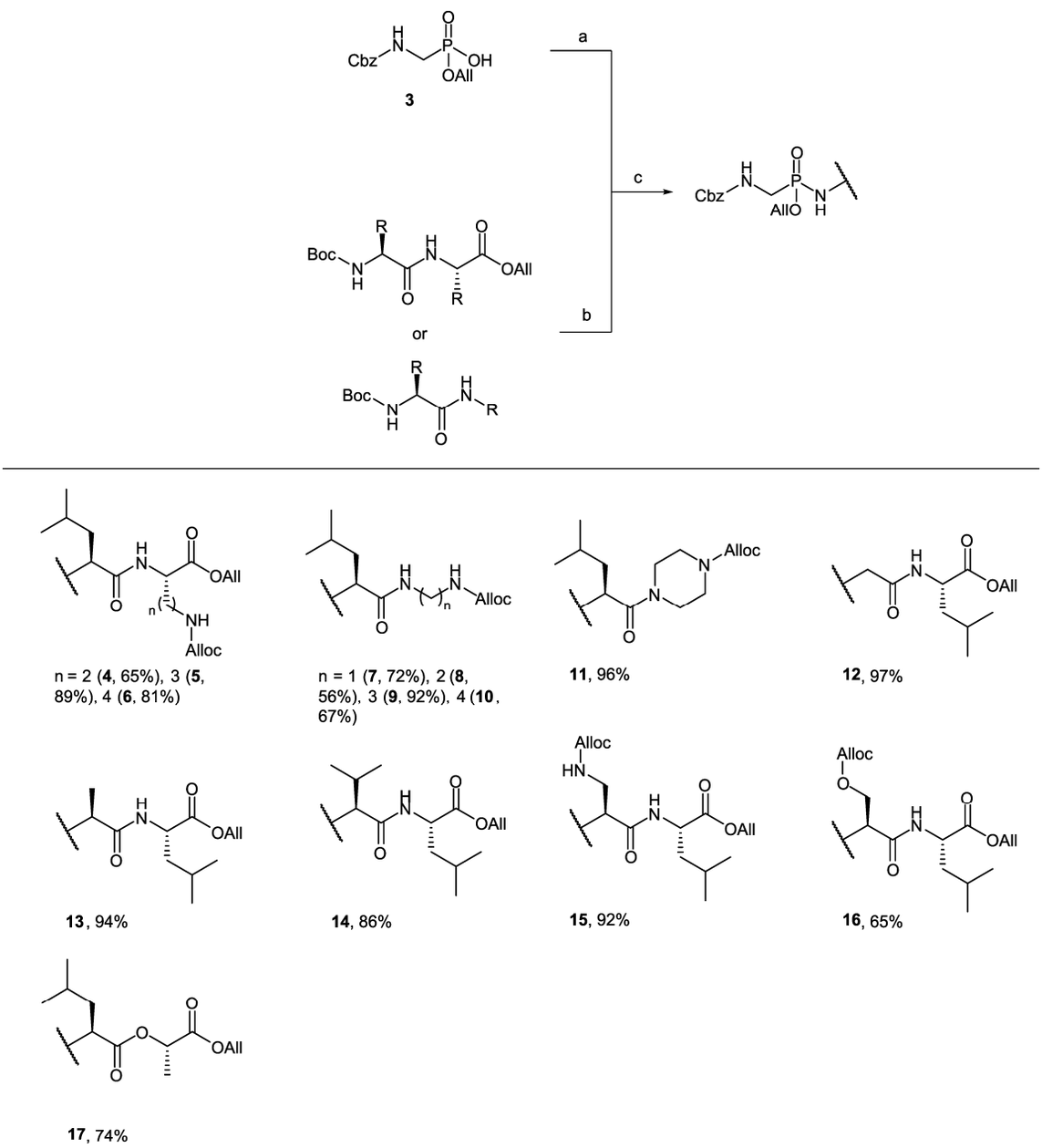
The synthesis of phosphonic acid diallylester **2** as precursor of the basic building block for the TLN inhibitor series was accomplished by a one-pot reaction of benzylcarbamate and paraformaldehyde in glacial acetic acid followed by an *Arbuzov* reaction with triallylphosphite. The resulting diallylester **2** was then selectively transferred to the monoallylester **3** by hydrolysis with aqueous NaOH (Scheme 2.2). Phosphonamidate coupling was achieved by reacting the corresponding phosphonochloridate with the free amino group of a peptidic building block (Scheme 2.3). The required phosphonochloridate was generated from **3** by activation with $SOCl_2$. As suggested by Hirschmann et al., the phosphonochloridate was treated with DIPEA prior to the addition of the coupling partner to generate an intermediate phosphonylammonium species with an increased reactivity towards nucleophiles [50]. Deallylation of the protected phosphonamidates was achieved by a palladium-catalyzed allyl transfer reaction using diethylamine as nucleophile. This method generated the fully deprotected inhibitor as a diethylammonium salt. After removal of the volatile side products, most hydrophobic contaminants were separated by a liquid-liquid extraction step. Analysis of the raw reaction products by qNMR indicated high conversion along with a purity of 70–80%, with excess diethylamine and catalyst residue as the major contaminants.

For compounds containing a primary amino group, the above procedure lead to the formation of non-stoichiometric mixtures of the inherently zwitterionic substances with diethylamine. Due to the rather close similarity of the pK_a values of diethylamine and the primary amino group, no consistent salt formation could be achieved. Additionally, the diethylammonium salts proved to be highly hygroscopic. Incubation of the raw reaction product with aqueous LiOH solution prior to liquid-liquid extraction lead to the formation of uniform lithium salts



1 **2** **3**

Scheme 2.2. Reagents and conditions: (a) $(CHO)_n$, AcOH, 50 °C, 4 h; (b) $P(OAll)_3$, neat, 110 °C, overnight (64%, two steps); (c) NaOH, H_2O, 1 h, 81%.



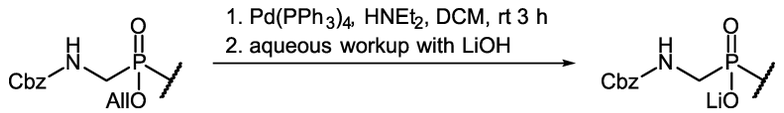
Scheme 2.3. Reagents and conditions: (a) $SOCl_2$, DCM, 0 °C to rt, 3 h; (b) 4 M HCl in dioxane, rt, 1 h; (c) DIPEA, DCM, 40 °C, overnight.

with greatly reduced hygroscopicity. By using a semi-preparative HPLC setup for the purification of functionalized phosphonopeptides, we frequently failed to achieve sufficient separation efficiencies, had trouble with detection and overall poor sample recovery. This might be attributed to the fact that the lability of the reaction product does not allow for the

addition of an acidic modifier such as TFA to the mobile phase, which, in consequence, severely deteriorates the peak shape. Additionally, there are reports that some phosphonamidates seem to show a tendency for decomposition during chromatographic purification, even without the addition of acidic modifiers [100]. Regularly, we also observed hydrolysis in preparative HPLC runs. As an alternative purification method, we investigated the use of solid-phase extraction (SPE). While predominantly used as a technique for sample preparation in analytical setups [101], SPE can be seen as a small-scale chromatography system that retains high sample concentration (narrow bands) and reduces exposure of the analyte to mobile and stationary phases to a minimum. While the separation efficiency is comparatively low, this technique has been successfully used for preparative purification of synthetic peptides, modified sugars and anthocyanidin-glucosides from plant material [102–104].

SPE cartridges with various phases and surface modifications are readily available from commercial suppliers, a number of which were tested for the purification of phosphonamidate peptides. The eluted fractions were analyzed by HPLC. Amino and cyano-modified silica gel phases and polymeric DPA-6S resin, however, did not show any retention of the desired compounds or contaminants in reversed phase mode. The use of a zwitterionic ammonium–sulfonic acid phase in HILIC mode was also investigated. This technique enabled the efficient separation of the desired products from apolar reaction side products. Unfortunately, direct application of alkaline samples to the column lead to hydrolysis of the surface modification. When the sample pH was adjusted, separation of the ionic product from buffer substances proved impossible. Efficient purification could be achieved with a C_{18}-modified silica gel phase. Here, the contaminants showed a strong retention, while the desired phosphonopeptides could be easily eluted with water or water/acetonitrile mixtures. By using these SPE phases, we were able to obtain the desired inhibitors with high purity. Table 2.1 shows a variety of phosphonopeptides prepared following the reported protocol. The scope of the procedure includes inhibitors featuring polar and basic functional groups. Additionally, the use of a palladium-catalyzed deprotection reaction grants access to compounds featuring a hydrolytically labile ester group, which have been challenging to prepare by traditional methods [105]. The novel SPE protocol enables a fast and efficient purification of the inhibitors while simultaneously reducing the amount of compound degradation in aqueous solvents. The purity of the isolated products is excellent and allows their use in biophysical assays such as isothermal titration calorimetry, a method that heavily relies on the accurate knowledge of ligand concentrations.

Table 2.1. Synthesis of peptidic phosphonamidate inhibitors.



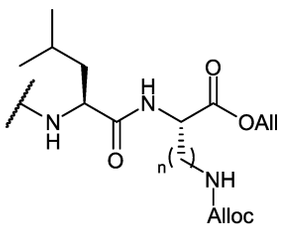
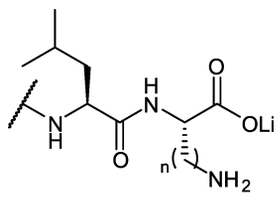
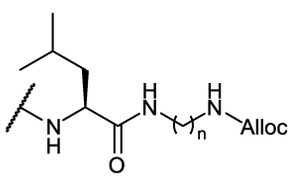
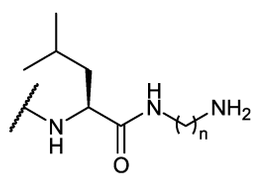
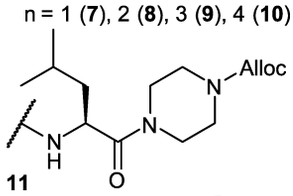
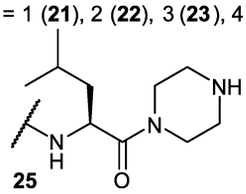
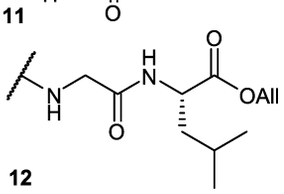
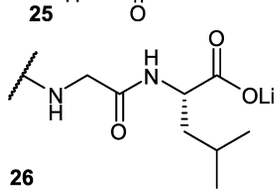
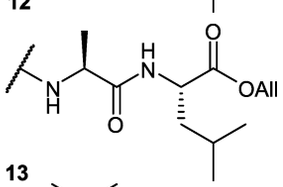
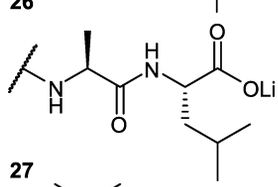
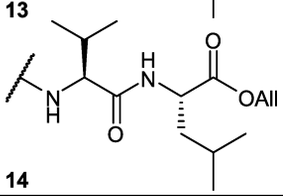
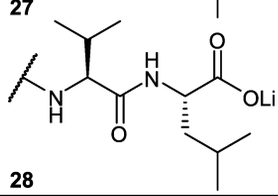
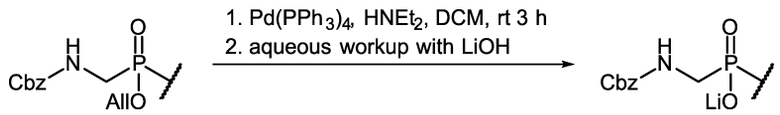
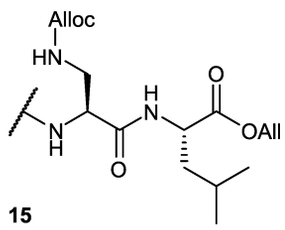
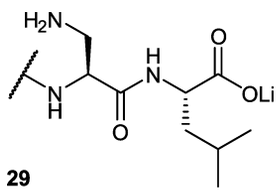
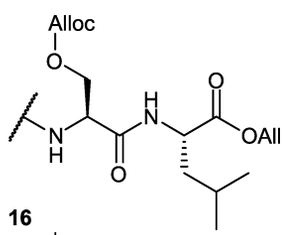
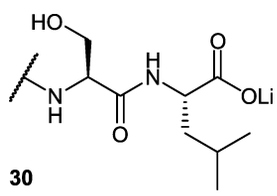
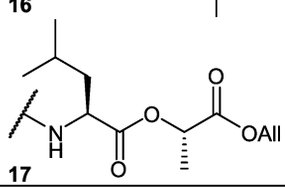
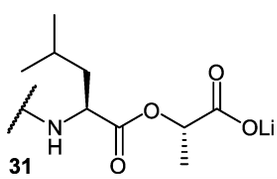
Entry	Reactant	Product	Yield	Purity (qNMR)
1	 n = 2 (**4**), 3 (**5**), 4 (**6**)	 n = 2 (**18**), 3 (**19**), 4 (**20**)	**18** = 73% **19** = 63% **20** = 66%	**18** = 95% **19** = 97% **20** = 99%
2	 n = 1 (**7**), 2 (**8**), 3 (**9**), 4 (**10**)	 n = 1 (**21**), 2 (**22**), 3 (**23**), 4 (**24**)	**21** = 78% **22** = 72% **23** = 72% **24** = 75%	**21** = 99% **22** = 99% **23** = 96% **24** = 96%
3	 **11**	 **25**	66%	97%
4	 **12**	 **26**	59%	97%
5	 **13**	 **27**	75%	97%
6	 **14**	 **28**	72%	96%

Table 2.1. Continued.



Entry	Reactant	Product	Yield	Purity (qNMR)
7	 **15**	 **29**	79%	96%
8	 **16**	 **30**	77%	97%
9	 **17**	 **31**	77%	97%

2.3 Experimental Section

^{1}H, ^{13}C and ^{31}P NMR spectra were recorded on a JEOL ECX-400 or JEOL ECA-500 instrument. All chemical shift values are reported in ppm relative to the non-deuterated solvent signal. An external standard was used for ^{31}P NMR spectra (85% H_3PO_4) and ^{13}C NMR spectra in D_2O (Trimethylsilyl propanoic acid). Quantitative NMR spectra were recorded using recommended instrument parameters for quantitative analysis (see Supporting Information). Relaxation delay was set to 50–60 s to ensure complete relaxation between pulses. Dimethyl terephthalate (Sigma-Aldrich, TraceCERT 99.99%) was used as an internal standard for quantification. All qNMR spectra were recorded in deuterated methanol. ESI-MS spectra were recorded on a Q-Trap 2000 system by Applied Biosystems. For high resolution ESI-MS, a LTQ-FT Ultra mass spectrometer (Thermo Fischer Scientific) was used. For HPLC chromatography, a Shimadzu LC-20 system equipped with a diode array detector was used. Analytic separations were carried out with a MN Nucleodur 100-5 C18 ec 4.6 × 250 mm column using a water-acetonitrile gradient with the addition of 0.1% TFA. While these conditions lead to hydrolysis of the analyte, the purity of the sample, regarding other contaminants, can be determined nevertheless. The absence of hydrolytic cleavage products in the sample was assured by ^{1}H and

^{31}P NMR. For SPE purification, Thermoscientific HyperSep C18 500 mg 3 mL cartridges were used. Melting points were determined with a Krüss KSP1N melting point meter.

2.4 Procedures

Benzyl ((bis(allyloxy)phosphoryl)methyl)carbamate (2): Benzyl-carbamate (756 mg, 5.00 mmol, 1.0 eq) and paraformaldehyde (180 mg, 6.00 mmol, 1.2 eq) were suspended in acetic acid and heated to 50 °C for 5 h. The solvent was removed under reduced pressure and triallyl phosphite (2.02 g, 10.00 mmol, 2.0 eq) was added to the residue. The mixture was heated to 110 °C for 14 h. All volatile components were removed under reduced pressure and the crude product was purified by silica gel chromatography (cyclohexane/EtOAc 1:1). The product was obtained as a colorless oil (1.042 g, 3.20 mmol, 64%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 3.66 (dd, *J* = 11.2, 6.0 Hz, 2H), 4.55 (dd, *J* = 7.6, 6.5 Hz, 4H), 5.05 (s, 1H), 5.11 (s, 2H), 5.24 (dd, *J* = 10.4, 1.1 Hz, 2H), 5.34 (dd, *J* = 17.1, 1.2 Hz, 2H), 5.84 – 5.99 (m, 2H), 7.28 – 7.41 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 36.2, 37.8, 67.0, 67.1, 67.4, 118.7, 118.7, 128.3, 128.4, 128.7, 132.6, 132.6, 136.2, 156.2. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 24.0. **MS (ESI)**: *m*/*z* $[M + H]^+$ calcd for $C_{15}H_{21}NO_5P$: 326.31; found: 326.08.

Benzyl (((allyloxy)(hydroxy)phosphoryl)methyl)carbamate (3): An aqueous solution of NaOH (80 mg, 2.00 mmol, 3.5 eq) was added to **2** (189 mg, 0.58 mmol, 1.0 eq) and the mixture was shaken vigorously until the solution cleared. After stirring for 1 h at rt, the mixture was washed with EtOAc (3 × 15 mL) and the organic phase was discarded. The aqueous phase was acidified to pH 1 and extracted with EtOAc (3 × 15 mL). The combined organic phases were washed with a saturated NaCl solution and dried over $MgSO_4$. The crude product was washed with hot TBME to give the product as a colorless solid (135 mg, 0.47 mmol, 81%); mp 91–92 °C. **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 3.50 – 3.70 (m, 2H), 4.45 – 4.57 (m, 2H), 5.10 (s, 2H), 5.22 (dd, *J* = 10.4, 0.9 Hz, 1H), 5.28 – 5.37 (m, 1H), 5.82 – 5.96 (m, 1H), 6.85 (bs, 1H), 7.27 – 7.38 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 66.9, 67.0, 67.5, 118.7, 128.3, 128.4, 128.7, 132.5, 132.5, 136.2, 156.5. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 25.2. **MS (ESI-):** *m*/*z* $[M - H]^-$ calcd for $C_{12}H_{15}NO_5P$: 284.23; found: 284.10.

General procedure for phosphonamidate coupling: $SOCl_2$ (4.0 eq) was dissolved in dry DCM under argon atmosphere and cooled to 0 °C. A solution of benzyl ((hydroxy(allyloxy)phosphoryl)methyl) carbamate (**3**, 2.0 eq) was added to the solution over 30 minutes. The cooling bath was removed and the reaction was stirred at rt for 4 h. All volatile components were removed under reduced pressure. In a second flask, the Boc-protected peptidic building block (1.0 eq, syntheses detailed in the Supporting Information) was treated

with HCl (4 M in dioxane, 4.0 eq) for 1 h and all volatile components were subsequently removed under reduced pressure. The activated phosphonic acid was dissolved in dry DCM. DIPEA (3.0 eq) and the deprotected peptide were consecutively added to the solution, after which the mixture was heated to 40 °C and stirred overnight. The solution was diluted with EtOAc and washed with 5% citric acid (3 × 10 mL), 1 M HCl (1 × 10 mL) and 1 M NaOH (3 × 10 mL). The separated organic phase was washed with brine and dried over $MgSO_4$. The product was obtained as a mixture of diastereomers and used in the next step without further purification.

Compound 4: Peptide **S3** (178 mg, 0.39 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (222 mg, 0.78 mmol), $SOCl_2$ (186 mg, 1.56 mmol), DIPEA (151 mg, 1.17 mmol). The product was obtained as a yellow oil (158 mg, 0.25 mmol, 65%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.84 – 1.03 (m, 6H), 1.41 – 1.54 (m, 1H), 1.57 – 1.92 (m, 3H), 1.99 – 2.22 (m, 1H), 2.99 – 3.16 (m, 1H), 3.32 – 3.48 (m, 1H), 3.51 – 3.99 (m, 4H), 4.41 – 4.76 (m, 6H), 5.03 – 5.42 (m, 8H), 5.75 – 5.99 (m, 3H), 7.25 – 7.40 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.8, 22.0, 23.0, 23.1, 24.61, 24.64, 32.2, 37.1, 44.0, 50.0, 53.9, 65.6, 65.68, 65.73, 65.8, 66.36, 66.43, 67.37, 67.41, 117.78, 117.83, 118.1, 118.4, 119.3, 119.4, 128.2, 128.32, 128.34, 128.37, 128.65, 128.67, 131.46, 131.52, 132.95, 133.00, 136.3, 136.4, 156.6, 156.7, 171.9, 172.1, 174.28, 174.32. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.6, 28.2. **MS (ESI+):** *m*/*z* $[M + H]^+$ calcd for $C_{29}H_{44}N_4O_9P$: 623.66; found: 623.26.

Compound 5: Peptide **S8** (235 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a yellow oil (283 mg, 0.44 mmol, 89%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.81 – 1.02 (m, 6H), 1.38 – 1.95 (m, 7H), 3.02 – 3.29 (m, 2H), 3.41 – 4.00 (m, 4H), 4.36 – 4.68 (m, 6H), 5.00 – 5.43 (m, 8H), 5.79 – 6.04 (m, 3H), 7.25 – 7.43 (m, 6H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.9, 22.0, 23.0, 23.0, 24.56, 24.59, 25.9, 26.1, 29.0, 38.0, 40.4, 43.88, 43.94, 52.2, 53.8, 65.6, 66.1, 66.2, 67.3, 67.4, 117.8, 118.1, 118.3, 119.1, 119.2, 128.2, 128.3, 128.6, 131.55, 131.58, 132.87, 132.94, 133.0, 133.1, 136.3, 156.7, 156.8, 171.8, 172.1, 173.9, 174.0. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.5, 28.4. **MS (ESI+):** *m*/*z* $[M + H]^+$ calcd for $C_{30}H_{46}N_4O_9P$: 637.69; found: 637.51.

Compound 6: Peptide **S9** (242 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a yellow oil (262 mg, 0.40 mmol, 81%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.83 – 1.06 (m, 6H), 1.28 – 2.11 (m, 9H), 3.02 – 3.25 (m, 2H), 3.40 – 3.96 (m, 4H), 4.40 – 4.76 (m, 6H), 5.02 – 5.45 (m, 8H), 5.76 – 6.03 (m, 3H), 6.91 – 7.04 (m, 1H), 7.26 – 7.49 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.8, 22.0, 22.3, 23.0, 23.1, 24.5, 29.3, 31.3, 38.0, 38.2, 39.4,

39.6, 40.2, 40.3, 43.7, 52.2, 52.4, 53.8, 65.6, 66.08, 66.13, 67.3, 117.7, 118.1, 118.4, 119.08, 119.12, 128.2, 128.3, 128.6, 131.6, 132.9, 133.0, 133.1, 136.3, 156.7, 172.0, 174.0. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.6, 28.3. **MS (ESI+):** *m*/*z* $[M + H]^+$ calcd for $C_{31}H_{48}N_4O_9P$: 651.72; found: 651.42.

Compound 7: Peptide **S11** (197 mg, 0.57 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (328 mg, 1.15 mmol), $SOCl_2$ (271 mg, 2.28 mmol), DIPEA (221 mg, 1.71 mmol). The product was obtained as a brown solid (185 mg, 0.36 mmol, 72%); mp 118–119 °C. **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.61 – 1.01 (m, 6H), 1.33 – 1.86 (m, 3H), 3.40 – 3.77 (m, 2H), 3.73 – 3.98 (m, 1H), 4.25 – 4.68 (m, 6H), 4.94 – 5.46 (m, 6H), 5.67 – 5.95 (m, 2H), 7.19 – 7.47 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 22.0, 22.1, 22.95, 23.01, 24.6, 24.7, 43.7, 46.4, 53.1, 53.4, 65.2, 65.9, 67.3, 117.9, 118.0, 128.26, 128.33, 128.6, 132.6, 132.9, 133.0, 136.36, 136.41, 156.8, 175.2, 175.4. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.6, 28.1. **MS (ESI+):** *m*/*z* $[M + H]^+$ calcd for $C_{23}H_{36}N_4O_7P$: 511.54; found: 511.48.

Compound 8: Peptide **S15** (179 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a colorless solid (146 mg, 0.28 mmol, 56%); mp 148–149 °C. **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.82 – 0.98 (m, 6H), 1.37 – 1.51 (m, 1H), 1.56 – 1.95 (m, 2H), 3.22 – 3.88 (m, 7H), 4.40 – 4.60 (m, 4H), 5.03 – 5.42 (m, 6H), 5.64 – 6.00 (m, 4H), 6.98 – 7.11 (m, 1H), 7.27 – 7.44 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.8, 22.1, 23.0, 23.1, 24.7, 38.0, 38.1, 39.86, 39.92, 40.7, 40.9, 43.7, 53.6, 53.8, 65.6, 65.8, 67.4, 67.5, 118.0, 118.3, 118.4, 128.3, 128.4, 128.7, 132.8, 133.0, 136.2, 156.9, 174.6. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.8, 28.4. **MS (ESI):** *m*/*z* $[M + H]^+$ calcd for $C_{24}H_{38}N_4O_7P$: 525.56; found: 525.26.

Compound 9: Peptide **S16** (334 mg, 0.90 mmol), HCl (4 M solution in dioxane, 1.0 mL, 4.00 mmol), **3** (513 mg, 1.80 mmol), $SOCl_2$ (428 mg, 3.60 mmol), DIPEA (349 mg, 2.70 mmol). The product was obtained as a yellow solid (446 mg, 0.83 mmol, 92%); mp 100–101 °C. **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.80 – 1.03 (m, 6H), 1.39 – 2.00 (m, 5H), 3.07 – 3.32 (m, 4H), 3.37 – 3.92 (m, 2H), 4.39 – 4.64 (m, 4H), 5.03 – 5.46 (m, 6H), 5.77 – 5.98 (m, 2H), 7.28 – 7.42 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.8, 22.0, 23.0, 23.1, 24.7, 29.7, 36.2, 36.3, 37.7, 37.8, 39.4, 39.6, 43.8, 53.4, 53.6, 65.5, 65.7, 67.3, 117.8, 118.1, 118.3, 128.3, 128.4, 128.7, 132.8, 133.0, 136.3, 156.9, 174.2, 174.5. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.9, 28.2. **MS (ESI):** *m*/*z* $[M + H]^+$ calcd for $C_{25}H_{40}N_4O_7P$: 539.59; found: 539.51.

Compound 10: Peptide **S17** (272 mg, 0.70 mmol), HCl (4 M solution in dioxane, 1.0 mL, 4.00 mmol), **3** (399 mg, 1.40 mmol), $SOCl_2$ (333 mg, 2.80 mmol), DIPEA (271 mg, 2.10 mmol). The product was obtained as a brown solid (258 mg, 0.47 mmol, 67%); mp 112–113 °C. **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.85 – 1.05 (m, 6H), 1.41 – 1.79 (m, 7H), 3.11 – 3.29 (m, 4H), 3.50 – 3.93 (m, 3H), 4.39 – 4.64 (m, 4H), 5.01 – 5.43 (m, 7H), 5.81 – 6.05 (m, 2H), 7.10 – 7.19 (m, 1H),

7.29 – 7.49 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.9, 22.1, 23.0, 23.1, 24.7, 26.5, 26.6, 27.3, 38.0, 38.2, 39.1, 40.6, 43.7, 53.3, 53.6, 65.4, 65.6, 67.3, 117.8, 118.1, 118.3, 128.3, 128.4, 128.7, 132.9, 133.1, 136.3, 156.6, 156.8, 173.8, 174.2. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.8, 28.2. **MS (ESI):** *m*/*z* $[M + NH_4]^+$ calcd for $C_{26}H_{45}N_5O_7P$: 570.31; found: 570.44.

Compound 11: Peptide **S19** (192 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a brown oil (264 mg, 0.48 mmol, 96%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.79 – 1.04 (m, 6H), 1.22 – 1.56 (m, 2H), 1.72 – 1.95 (m, 1H), 3.22 – 3.81 (m, 10H), 4.23 – 4.37 (m, 1H), 4.40 – 4.68 (m, 4H), 5.00 – 5.42 (m, 6H), 5.78 – 6.05 (m, 2H), 7.19 – 7.44 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.4, 21.7, 23.4, 23.5, 24.4, 24.6, 38.3, 39.3, 42.1, 42.2, 43.6, 43.9, 45.3, 49.1, 49.6, 53.5, 65.0, 65.2, 66.5, 67.2, 67.3, 117.9, 118.0, 128.2, 128.3, 128.59, 128.64, 132.7, 132.9, 133.0, 136.4, 154.9, 155.0, 156.4, 172.6, 173.2. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.4, 28.0. **MS (ESI+):** *m*/*z* $[M + NH_4]^+$ calcd for $C_{26}H_{43}N_5O_7P$: 568.63; found: 568.44.

Compound 12: Peptide **S21** (164 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a brown oil (240 mg, 0.48 mmol, 97%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.78 – 0.98 (m, 6H), 1.54 – 1.75 (m, 3H), 3.32 – 3.52 (m, 2H), 3.70 – 3.97 (m, 2H), 4.29 – 4.43 (m, 2H), 4.47 – 4.71 (m, 3H), 4.98 – 5.45 (m, 6H), 5.75 – 6.03 (m, 2H), 7.23 – 7.46 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.8, 22.8, 24.8, 37.0, 38.4, 40.6, 51.4, 66.1, 66.8, 116.8, 118.8, 128.0, 128.1, 128.5, 131.7, 134.4, 134.5, 136.8, 156.8, 156.9, 166.9, 172.9. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 19.3. **MS (ESI+):** *m*/*z* $[M + NH_4]^+$ calcd for $C_{23}H_{38}N_4O_7P$: 513.55; found: 513.41.

Compound 13: Peptide **S22** (171 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a brown oil (239 mg, 0.47 mmol, 94%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.86 – 1.04 (m, 6H), 1.23 – 1.49 (m, 3H), 1.53 – 1.82 (m, 3H), 3.45 – 3.81 (m, 2H), 3.84 – 4.05 (m, 1H), 4.44 – 4.73 (m, 5H), 5.09 (d, *J* = 10.8 Hz, 2H), 5.19 – 5.47 (m, 4H), 5.71 (dd, *J* = 28.8, 23.4 Hz, 1H), 5.80 – 6.01 (m, 2H), 7.28 – 7.44 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.4, 21.6, 21.9, 23.0, 25.0, 37.7, 38.1, 39.1, 39.5, 41.09, 41.13, 50.6, 50.9, 51.0, 51.1, 65.25, 65.31, 65.37, 65.43, 66.0, 67.3, 118.2, 118.3, 118.9, 128.2, 128.26, 128.31, 128.6, 131.7, 132.80, 132.85, 132.90, 136.4, 156.7, 156.8, 172.7, 173.0, 173.7, 173.9. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.4, 27.6. **MS (ESI+):** *m*/*z* $[M + NH_4]^+$ calcd for $C_{24}H_{40}N_4O_7P$: 527.58; found: 527.43.

Compound 14: Peptide **S23** (185 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a brown oil (231 mg, 0.43 mmol, 86%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.82 – 1.01 (m, 9H), 1.01 – 1.15 (m, 3H), 1.59 – 1.83 (m, 3H), 2.04 – 2.43 (m, 1H), 3.48 –

3.84 (m, 4H), 4.43 – 4.69 (m, 4H), 5.05 – 5.15 (m, 2H), 5.17 – 5.39 (m, 4H), 5.79 – 6.01 (m, 2H), 7.27 – 7.42 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 17.5, 17.6, 19.3, 19.4, 21.6, 21.7, 22.9, 23.0, 24.9, 24.9, 32.1, 32.2, 37.8, 38.1, 39.3, 39.5, 40.9, 41.0, 50.9, 51.1, 60.0, 60.6, 65.2, 65.3, 65.4, 65.97, 66.01, 67.17, 67.23, 117.9, 118.0, 118.85, 118.88, 128.1, 128.2, 128.55, 128.57, 131.7, 132.80, 132.87, 132.90, 133.0, 136.4, 156.55, 156.60, 156.7, 156.8, 172.6, 172.9, 173.0, 173.1. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.9, 28.3. **MS (ESI+):** *m*/*z* $[M + NH_4]^+$ calcd for $C_{26}H_{44}N_4O_7P$: 555.63; found: 555.37.

Compound 15: Peptide **S24** (221 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a brown oil (280 mg, 0.46 mmol, 92%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.83 – 1.01 (m, 6H), 1.58 – 1.77 (m, 3H), 3.37 – 3.74 (m, 3H), 4.10 (m, 2H), 4.37 – 4.70 (m, 7H), 5.02 – 5.37 (m, 8H), 5.81 – 5.98 (m, 3H), 7.27 – 7.41 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.6, 23.0, 24.9, 38.1, 39.6, 40.4, 45.2, 45.4, 51.3, 51.3, 54.8, 55.4, 65.2, 65.3, 65.5, 65.5, 65.97, 66.03, 66.1, 67.2, 67.3, 117.8, 118.2, 118.3, 118.9, 128.16, 128.18, 128.3, 128.55, 128.58, 131.7, 132.8, 136.3, 136.4, 156.7, 157.5, 171.3, 171.5, 172.6, 172.7. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.6. **MS (ESI+):** *m*/*z* $[M + NH_4]^+$ calcd for $C_{28}H_{45}N_5O_9P$: 626.67; found: 626.38.

Compound 16: Peptide **S26** (221 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a brown oil (199 mg, 0.33 mmol, 65%). **^{1}H NMR** ($CDCl_3$, 500 MHz): δ = 0.85 – 1.03 (m, 6H), 1.57 – 1.80 (m, 3H), 3.58 – 3.79 (m, 2H), 4.11 – 4.46 (m, 3H), 4.49 – 4.70 (m, 7H), 5.04 – 5.18 (m, 2H), 5.18 – 5.40 (m, 6H), 5.81 – 5.97 (m, 3H), 7.28 – 7.41 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 21.8, 23.0, 24.9, 40.9, 41.1, 51.2, 51.3, 54.2, 54.2, 65.6, 66.1, 66.2, 67.5, 67.5, 69.0, 118.5, 118.9, 119.0, 119.4, 119.4, 128.3, 128.4, 128.7, 131.4, 131.7, 132.7, 132.7, 136.3, 154.6, 156.8, 169.6, 172.5. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 27.1, 27.7. **MS (ESI+):** *m*/*z* $[M + NH_4]^+$ calcd for $C_{28}H_{44}N_4O_{10}P$: 627.65; found: 627.33.

Compound 17: Peptide **S28** (172 mg, 0.50 mmol), HCl (4 M solution in dioxane, 0.5 mL, 2.00 mmol), **3** (285 mg, 1.00 mmol), $SOCl_2$ (238 mg, 2.00 mmol), DIPEA (194 mg, 1.50 mmol). The product was obtained as a yellow oil (189 mg, 0.37 mmol, 74%). **^{1}H NMR** ($CDCl_3$, 400 MHz): δ = 0.86 – 1.04 (m, 6H), 1.42 – 1.64 (m, 4H), 1.65 – 1.91 (m, 2H), 2.92 – 3.17 (m, 1H), 3.38 – 3.85 (m, 2H), 3.98 – 4.17 (m, 1H), 4.39 – 4.73 (m, 4H), 5.01 – 5.39 (m, 7H), 5.76 – 6.02 (m, 2H), 7.28 – 7.44 (m, 5H). **^{13}C NMR** ($CDCl_3$, 101 MHz): δ = 16.87, 16.90, 21.3, 21.5, 23.0, 23.1, 24.5, 24.6, 37.9, 38.1, 39.3, 39.6, 43.17, 43.23, 43.58, 43.64, 52.2, 52.3, 65.0, 65.1, 65.2, 65.3, 66.0, 66.1, 67.20, 67.24, 69.3, 69.4, 117.8, 118.0, 119.03, 119.06, 128.1, 128.2, 128.3, 128.57, 128.59, 131.38, 131.42, 132.9, 133.0, 136.3, 136.4, 156.4, 170.2, 174.0, 174.5. **^{31}P NMR** ($CDCl_3$, 162 MHz): δ = 26.6, 27.5. **MS (ESI+):** *m*/*z* $[M + Na]^+$ calcd for $C_{24}H_{35}N_2NaO_8P$: 533.51; found: 533.34.

General procedure for phosphonamidate deprotection: The protected phosphonamidate (1.0 eq) was dissolved in DCM and diethylamine (20.0 eq) and $Pd(PPh_3)_4$ (10 mol%) were consecutively added to the solution. After stirring at rt for 3 h, all volatile components were removed under reduced pressure. The residue was taken up in DCM and the solution was extracted with water containing 2.2 eq LiOH (3 × 2 mL). The combined aqueous phases were washed with DCM and lyophilized. The crude product was purified using a reversed phase SPE cartridge (500 mg, C18). The SPE column was conditioned with 1 mL MeCN followed by 5 mL H_2O. The sample was dissolved in 1 mL H_2O, slowly passed through the column and eluted using H_2O or H_2O/MeCN mixtures. The eluted fractions were analyzed by HPLC. All fractions containing the desired product in sufficiently pure form were combined and lyophilized. The purity of the products was analyzed by quantitative NMR spectroscopy.

Compound 18: Compound **4** (93 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid (57 mg, 95% purity by qNMR, 0.11 mmol, 73%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.82 – 0.94 (m, 6H), 1.35 – 1.50 (m, 2H), 1.56 – 1.68 (m, 1H), 1.85 – 1.94 (m, 1H), 1.99 – 2.07 (m, 1H), 2.78 – 2.89 (m, 2H), 3.21 – 3.35 (m, 2H), 3.62 – 3.67 (m, 1H), 4.16 (dd, *J* = 8.4, 5.0 Hz, 1H), 5.05 – 5.19 (m, 2H), 7.34 – 7.43 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 21.2, 22.2, 24.0, 36.8, 39.7 (d, J_{CP} = 135.7 Hz), 43.21, 43.27, 52.5, 54.0, 67.1, 127.8, 128.4, 128.8, 136.4, 158.2, 177.4, 177.7. **^{31}P NMR** (D_2O, 162 MHz): δ = 18.2. **HRMS (ESI):** *m*/*z* $[M + H]^+$ calcd for $C_{19}H_{32}N_4O_7P$: 459.2003; found: 459.2003.

Compound 19: Compound **5** (96 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid (51 mg, 97% purity by qNMR, 0.10 mmol, 63%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.80 – 1.09 (m, 6H), 1.44 – 1.98 (m, 7H), 2.73 – 3.00 (m, 2H), 3.19 – 3.57 (m, 2H), 3.67 – 3.87 (m, 1H), 4.15 – 4.33 (m, 1H), 5.06 – 5.42 (m, 2H), 7.38 – 7.75 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 21.2, 22.2, 24.0, 25.0, 28.8, 39.5, 39.7 (d, J_{CP} = 136.1 Hz), 43.2, 54.0, 54.4, 67.1, 127.8, 128.4, 128.8, 138.0, 158.3, 177.4, 178.3. **^{31}P NMR** (D_2O, 202 MHz): δ = 18.2. **HRMS (ESI):** *m*/*z* $[M + H]^+$ calcd for $C_{20}H_{34}N_4O_7P$: 473.2160; found: 473.2161.

Compound 20: Compound **6** (100 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid (50 mg, 99% purity by qNMR, 0.10 mmol, 66%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 500 MHz): δ = 0.91 – 1.02 (m, 6H), 1.33 – 1.53 (m, 5H), 1.73 – 1.84 (m, 4H), 2.72 – 2.81 (m, 2H), 3.30 – 3.38 (m, 2H), 3.74 – 3.76 (m, 1H), 4.22 (dd, *J* = 7.2, 5.3 Hz, 1H), 5.17 – 5.23 (m, 2H), 7.47 – 7.54 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 21.3, 21.8, 22.2, 24.0, 26.4, 31.1, 39.2, 39.7 (d, J_{CP} = 136.0 Hz), 43.2, 54.0, 54.6, 67.1, 127.7, 128.3, 128.8, 136.5, 158.3, 177.2, 178.5. **^{31}P**

NMR (D_2O, 202 MHz): δ = 18.2. **HRMS (ESI):** *m*/*z* $[M + H]^+$ calcd for $C_{21}H_{36}N_4O_7P$: 487.2316; found: 487.2315.

Compound 21: Compound **7** (77 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid (39 mg, 99% purity by qNMR, 0.12 mmol, 78%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.75 – 0.90 (m, 6H), 1.31 – 1.47 (m, 2H), 1.55 – 1.68 (m, 1H), 3.17 – 3.34 (m, 2H), 3.51 – 3.61 (m, 1H), 3.92 (s, 2H), 5.01 – 5.16 (m, 2H), 7.29 – 7.49 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 21.0, 22.3, 24.0, 39.7 (d, J_{CP} = 137.0 Hz), 43.0, 47.9, 54.1, 67.1, 127.7, 128.4, 128.8, 136.4, 158.3, 178.5. **^{31}P NMR** (D_2O, 162 MHz): δ = 17.9. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{16}H_{27}N_4O_5PNa$: 409.1611; found: 409.1614.

Compound 22: Compound **8** (146 mg, 0.28 mmol), $Pd(PPh_3)_4$ (32 mg, 0.03 mmol), $HNEt_2$ (404 mg, 5.60 mmol), LiOH (26 mg, 0.62 mmol). The product was obtained as a colorless solid (82 mg, 99% purity by qNMR, 0.09 mmol, 72%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.74 – 0.94 (m, 6H), 1.33 – 1.50 (m, 2H), 1.52 – 1.71 (m, 1H), 2.69 – 2.90 (m, 2H), 3.12 – 3.43 (m, 4H), 3.53 – 3.73 (m, 1H), 4.98 – 5.27 (m, 2H), 7.29 – 7.55 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 21.1, 22.3, 24.1, 38.8, 39.42 (d, J_{CP} = 137.9 Hz), 39.46, 42.9, 54.4, 67.1, 127.8, 128.4, 128.8, 136.4, 158.2, 179.0. **^{31}P NMR** (D_2O, 162 MHz): δ = 18.4. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{17}H_{29}N_4O_5PNa$: 423.1768; found: 423.1766.

Compound 23: Compound **9** (54 mg, 0.10 mmol), $Pd(PPh_3)_4$ (12 mg, 0.01 mmol), $HNEt_2$ (146 mg, 2.00 mmol), LiOH (9 mg, 0.22 mmol). The product was obtained as a colorless solid (28 mg, 92% purity by qNMR, 0.05 mmol, 55%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.65 – 0.98 (m, 6H), 1.29 – 1.52 (m, 2H), 1.51 – 1.73 (m, 3H), 2.55 – 2.76 (m, 2H), 3.01 – 3.38 (m, 4H), 3.46 – 3.67 (m, 1H), 4.98 – 5.22 (m, 2H), 7.24 – 7.53 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 21.1, 22.2, 24.0, 29.3, 36.4, 37.5, 40.0 (d, J_{CP} = 136.2 Hz), 43.2, 54.2, 67.1, 127.8, 128.4, 128.8, 136.4, 158.3, 178.4. **^{31}P NMR** (D_2O, 162 MHz): δ = 18.1. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{18}H_{31}N_4O_5PNa$: 437.1924; found: 437.1923.

Compound 24: Compound **10** (50 mg, 0.09 mmol), $Pd(PPh_3)_4$ (10 mg, 0.01 mmol), $HNEt_2$ (132 mg, 1.80 mmol), LiOH (8 mg, 0.18 mmol). The product was obtained as a colorless solid (30 mg, 96% purity by qNMR, 0.07 mmol, 75%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.85 – 1.01 (m, 6H), 1.38 – 1.61 (m, 6H), 1.61 – 1.76 (m, 1H), 2.61 – 2.73 (m, 2H), 3.15 – 3.25 (m, 2H), 3.26 – 3.47 (m, 2H), 3.61 – 3.76 (m, 1H), 5.19 (s, 2H), 7.38 – 7.60 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 21.2, 22.2, 24.0, 25.7, 27.6, 38.8, 39.0, 39.7 (d, J_{CP} = 136.4 Hz), 43.3, 54.1, 67.1, 127.8, 128.4, 128.8, 136.4, 158.3, 178.0. **^{31}P NMR** (D_2O, 162 MHz): δ = 17.9. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{19}H_{33}N_4O_5PNa$: 451.2081; found: 451.2084.

Compound 25: Compound **11** (44 mg, 0.08 mmol), $Pd(PPh_3)_4$ (9 mg, 0.01 mmol), $HNEt_2$ (117 mg, 1.60 mmol), LiOH (8 mg, 0.18 mmol). The product was obtained as a colorless solid (24 mg, 97% purity by qNMR, 0.05 mmol, 66%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.81 – 0.91 (m, 6H), 1.22 – 1.42 (m, 2H), 1.57 – 1.76 (m, 1H), 2.61 – 2.93 (m, 4H), 3.07 – 3.24 (m, 2H), 3.27 – 3.63 (m, 4H), 4.06 – 4.18 (m, 1H), 5.03 – 5.14 (m, 2H), 7.31 – 7.46 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 21.2, 22.5, 24.0, 39.7 (d, J_{CP} = 135.2 Hz), 42.6, 43.5, 44.1, 44.5, 46.0, 49.2, 67.1, 127.9, 128.4, 128.8, 136.4, 158.1, 175.4. **^{31}P NMR** (D_2O, 162 MHz): δ = 18.0. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{19}H_{31}N_4O_5PNa$: 449.1924; found: 449.1922.

Compound 26: Compound **12** (74 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid (39 mg, 97% purity by qNMR, 0.09 mmol, 59%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 500 MHz): δ = 0.91 – 1.04 (m, 6H), 1.63 – 1.75 (m, 3H), 3.35 – 3.46 (m, 2H), 3.51 – 3.63 (m, 2H), 4.26 – 4.35 (m, 1H), 5.17 – 5.28 (m, 2H), 7.44 – 7.58 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 20.9, 22.4, 24.6, 38.4, 39.2 (d, J_{CP} = 135.2 Hz), 44.3, 53.7, 67.1, 127.87, 128.4, 128.8, 136.5, 158.2, 174.2, 179.9. **^{31}P NMR** (D_2O, 202 MHz): δ = 20.2. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{17}H_{26}N_3O_7PNa$: 483.1401; found: 483.1400.

Compound 27: Compound **13** (76 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid (50 mg, 97% purity by qNMR, 0.11 mmol, 75%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 500 MHz): δ = 0.93 (d, *J* = 6.2 Hz, 3H), 0.98 (d, *J* = 6.1 Hz, 3H), 1.33 (d, *J* = 7.1 Hz, 3H), 1.61 – 1.73 (m, 3H), 3.20 – 3.44 (m, 2H), 3.77 – 3.83 (m, 1H), 4.24 – 4.28 (m, 1H), 5.16 – 5.25 (m, 2H), 7.46 – 7.57 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 20.3, 20.9, 22.5, 24.6, 39.4 (d, J_{CP} = 136.1 Hz), 40.8, 50.9, 53.8, 67.1, 127.8, 128.4, 128.8, 136.5, 158.3, 177.1, 179.9. **^{31}P NMR** (D_2O, 202 MHz): δ = 18.9. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{18}H_{28}N_3O_7PNa$: 452.1557; found: 452.1555.

Compound 28: Compound **14** (81 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid (52 mg, 96% purity by qNMR, 0.11 mmol, 72%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.71 – 0.95 (m, 12H), 1.46 – 1.63 (m, 3H), 1.91 – 2.05 (m, 1H), 3.10 – 3.34 (m, 2H), 3.39 – 3.47 (m, 1H), 4.12 – 4.23 (m, 1H), 5.02 – 5.20 (m, 2H), 7.32 – 7.47 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 16.8, 18.8, 20.9, 22.4, 24.6, 31.5, 39.3 (d, J_{CP} = 136.8 Hz), 40.9, 53.9, 60.8, 67.1, 127.8, 128.4, 128.8, 136.5, 158.3, 176.2, 179.8. **^{31}P NMR** (D_2O, 162 MHz): δ = 18.5. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{20}H_{32}N_3O_7PNa$: 480.187; found: 480.1868.

Compound 29: Compound **15** (91 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid

(56 mg, 96% purity by qNMR, 0.12 mmol, 79%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 500 MHz): δ = 0.89 – 1.04 (m, 6H), 1.61 – 1.75 (m, 3H), 2.84 – 3.08 (m, 2H), 3.28 – 3.52 (m, 2H), 3.78 – 3.90 (m, 1H), 4.20 – 4.29 (m, 1H), 5.13 – 5.34 (m, 2H), 7.39 – 7.64 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 20.8, 22.5, 24.7, 39.5 (d, J_{CP} = 134.7 Hz), 40.5, 44.4, 54.1, 56.0, 67.1, 127.8, 128.4, 128.8, 136.5, 158.2, 173.9, 179.8. **^{31}P NMR** (D_2O, 202 MHz): δ = 18.7. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{18}H_{29}N_4O_7PNa$: 467.1666; found: 467.1665.

Compound 30: Compound **16** (91 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid (55 mg, 97% purity by qNMR, 0.11 mmol, 77%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.70 – 0.92 (m, 6H), 1.48 – 1.68 (m, 3H), 3.15 – 3.42 (m, 2H), 3.55 – 3.80 (m, 3H), 4.10 – 4.22 (m, 1H), 5.01 – 5.20 (m, 2H), 7.23 – 7.49 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 20.9, 22.4, 24.6, 39.5 (d, J_{CP} = 136.7 Hz), 40.7, 54.1, 56.8, 63.8, 67.1, 127.8, 128.4, 128.8, 136.5, 158.2, 174.0, 179.9. **^{31}P NMR** (D_2O, 162 MHz): δ = 18.9. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{18}H_{28}N_3O_8PNa$: 468.1506; found: 468.1504.

Compound 31: Compound **17** (77 mg, 0.15 mmol), $Pd(PPh_3)_4$ (17 mg, 0.02 mmol), $HNEt_2$ (219 mg, 3.00 mmol), LiOH (14 mg, 0.33 mmol). The product was obtained as a colorless solid (53 mg, 97% purity by qNMR, 0.12 mmol, 77%); decomposition above 200 °C. **^{1}H NMR** (D_2O, 400 MHz): δ = 0.81 – 0.98 (m, 6H), 1.33 – 1.47 (m, 4H), 1.53 – 1.82 (m, 2H), 3.08 – 3.43 (m, 2H), 3.74 – 3.91 (m, 1H), 4.81 – 4.85 (m, 1H), 5.05 – 5.24 (m, 2H), 7.20 – 7.58 (m, 5H). **^{13}C NMR** (D_2O, 101 MHz): δ = 16.8, 21.0, 22.4, 23.9, 39.7 (d, J_{CP} = 136.1 Hz), 42.8, 52.8, 67.1, 72.7, 127.8, 128.4, 128.8, 136.4, 158.3, 177.3, 178.6. **^{31}P NMR** (D_2O, 202 MHz): δ = 18.3. **HRMS (ESI):** *m*/*z* $[M + Na]^+$ calcd for $C_{18}H_{27}N_2O_8PNa$: 453.1397; found: 453.1395.

2.5 Acknowledgement

This work was funded by the European Research Council (ERC) of the European Union (grant 268145-DrugProfilBind).

2.6 Supporting Information

2.6.1 General procedures

General procedure for the synthesis of allyl esters

The *N*-protected amino acid (1.0 eq) was dissolved in DMF. Cs_2CO_3 (1.0 eq) and allyl bromide (1.3 eq) were added and the mixture was stirred at rt overnight. The solvent was removed and

the resulting residue was taken up in EtOAc. The solution was extracted with 1 M HCl (3 x 10 mL) and sat. $NaHCO_3$ (3 x 10 mL), washed with brine and dried over $MgSO_4$.

General procedure for peptide coupling

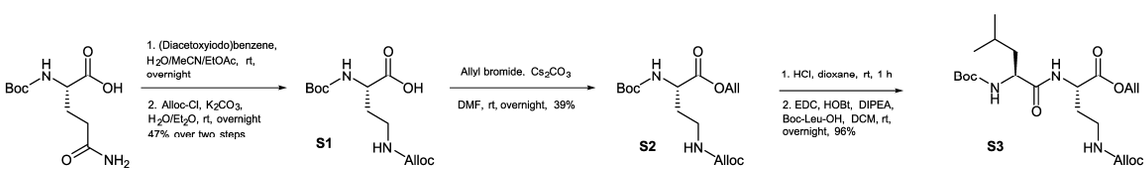
The respective Boc-protected amine was dissolved in 2 mL of DCM and HCl (4 M solution in dioxane, 2.0 eq) was added. The mixture was stirred for 2 h and the solvent was removed under reduced pressure. A carboxylic acid (1.0 eq), EDC (1.3 eq), HOBt (1.3 eq) and DIPEA (2.5 eq) were added to the flask. The mixture was dissolved in DCM and the stirred at rt overnight. The solution was diluted with EtOAc and extracted with 1 M HCl (3 x 15 mL) and sat. $NaHCO_3$ (3 x 15 mL). The combined organic phases were washed with brine and dried over $MgSO_4$. The crude product was purified by silica gel column chromatography (DCM/MeOH 50:1).

General procedure for the synthesis of protected diamines

The respective Boc-protected diamine and NEt_3 (2.0 eq) were dissolved in DCM. Alloc-Cl (1.2 eq) was added dropwise under ice cooling. The reaction mixture was stirred at rt overnight. After removal of the solvent, the residue was taken up in EtOAc and extracted with 1 M HCl (3 x 10 mL), sat. $NaHCO_3$ solution (3 x 10 mL) and brine (1 x 10 mL). The organic phase was dried over $MgSO_4$ and the raw product was purified by silica gel chromatography (DCM/MeOH 50:1).

2.6.2 Synthesis of peptidic precursors

Synthesis of Peptide S3



Scheme S2.1. Synthesis of peptide **S3**.

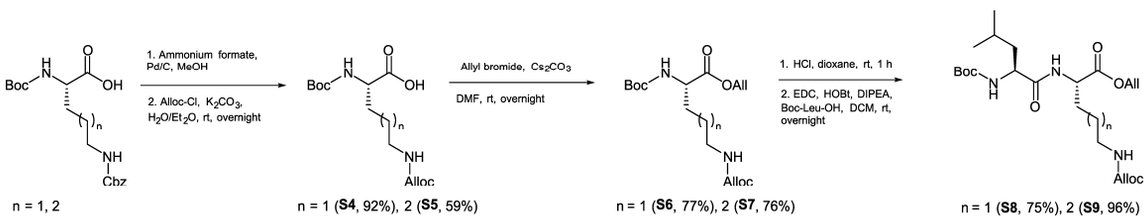
*Boc-Dab(Alloc)-OH (***S1***):* Boc-Gln-OH (1.72 g, 7.00 mmol, 1.0 eq) and (Diacetoxyiodo)benzene (2.71 g, 7.40 mmol, 1.20 eq) were dissolved in an ice cold mixture of EtOAc/MeCN/H_2O (2:2:1). The mixture was stirred at 50 °C overnight and afterwards extracted with EtOAc (3 x 10 mL). K_2CO_3 (2.42 g, 17.50 mmol, 2.5 eq) was added to the aqueous phase. A solution of Alloc-Cl (1.01 g, 8.40 mmol, 1.2 eq) in Et_2O was added to the flask and the

biphasic mixture was stirred at rt overnight. The phases were separated and the aqueous phase was washed with EtOAc (3 x 10 mL) and the organic phases were discarded. The pH was adjusted to 1 and the resulting solution was extracted with EtOAc (3 x 15 mL). The combined organic phases were washed with brine and dried over $MgSO_4$. The product was obtained as a yellow oil (985 mg, 3.26 mmol, 47%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.45 (s, 9H), 1.77 – 2.13 (m, 2H), 3.03 – 3.58 (m, 2H), 4.30 – 4.47 (m, 1H), 4.52 – 4.68 (m, 2H), 5.21 – 5.43 (m, 3H), 5.46 – 5.58 (m, 1H), 5.83 – 6.03 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 28.4, 42.7, 53.6, 54.6, 66.3, 81.0, 118.1, 132.6, 156.5, 157.5, 173.6. **MS (ESI-)** *m/z* calculated for $C_{13}H_{21}N_2O_6$ $[M-H]^-$: 301.32; found 301.25.

*Boc-Dab(Alloc)-OAll (***S2***):* Synthesized after general procedure 1.1. Boc-Dab(Alloc)-OH (**S2**, 537 mg, 1.77 mmol, 1.0 eq) was dissolved in MeCN. DIPEA (458 mg, 3.54 mmol, 2.0 eq) and allyl bromide (645 mg, 5.33 mmol, 3.0 eq) were added and the mixture was stirred at 40 °C overnight. The solvent was removed and the resulting residue was taken up in EtOAc. The solution was extracted with 1 M HCl (3 x 10 mL) and sat. $NaHCO_3$ (3 x 10 mL), washed with brine and dried over $MgSO_4$. The crude product was purified by silica gel chromatography (cyclohexanes/EtOAc 2:1). The product was obtained as a yellow oil (235 mg, 0.69 mmol, 39%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 1.44 (s, 9H), 1.64 – 1.72 (m, 1H), 2.05 – 2.14 (m, 1H), 3.01 – 3.11 (m, 1H), 3.47 – 3.54 (m, 1H), 4.36 – 4.42 (m, 1H), 4.52 – 4.61 (m, 2H), 4.63 (d, *J* = 5.8 Hz, 2H), 5.18 – 5.37 (m, 5H), 5.48 (bs, 1H), 5.85 – 5.98 (m, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 28.4, 33.8, 37.2, 51.1, 65.7, 66.3, 80.4, 117.7, 119.3, 131.5, 133.1, 156.0, 156.4, 172.4. **MS (ESI+)** *m/z* calculated for $C_{16}H_{30}N_3O_6$ $[M+NH_4]^+$: 360.21; found 360.19.

Boc-Leu-Dab(Alloc)-OAll (**S3**): Synthesized after general procedure 1.2. Boc-Dab(Alloc)-OAll (**S4**, 685 mg, 2.00 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Leu-OH (463 mg, 2.00 mmol), EDC (498 mg, 2.60 mmol), HOBt (351 mg, 2.60 mmol) and DIPEA (646 mg, 5.00 mmol). The product was obtained as a yellow oil (874 mg, 1.91 mmol, 96%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.94 (t, *J* = 6.1 Hz, 6H), 1.43 (s, 9H), 1.58 – 1.76 (m, 4H), 2.10 – 2.21 (m, 1H), 2.91 – 3.03 (m, 1H), 3.45 – 3.56 (m, 1H), 4.03 – 4.13 (m, 1H), 4.50 – 4.72 (m, 5H), 4.87 (d, *J* = 7.1 Hz, 1H), 5.16 – 5.37 (m, 4H), 5.67 (bs, 1H), 5.82 – 5.99 (m, 2H), 6.77 (d, *J* = 7.7 Hz, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.1, 23.0, 24.8, 28.3, 33.4, 36.7, 41.1, 49.7, 53.6, 65.6, 66.4, 80.4, 117.6, 119.4, 131.3, 133.1, 155.8, 156.5, 171.9, 173.4. **MS (ESI+)** *m/z* calculated for $C_{22}H_{41}N_4O_7$ $[M+NH_4]^+$: 473.30; found 473.26.

Synthesis of Peptides S8–S9



Scheme S2.2. Synthesis of peptides **S8–S9**.

*Boc-Orn(Alloc)-OH (***S4***):* Boc-Orn(Cbz)-OH (3.66 g, 10.00 mmol, 1.0 eq) and ammonium formate (3.15 g, 50.00 mmol, 5.0 eq) were dissolved in MeOH. Palladium on activated charcoal (10% Pd, 851 mg, 0.80 mmol, 8 mol%) was added and the resulting suspension was stirred for 2 h at rt. The mixture was filtered over Celite, the solvent was removed under reduced pressure. The residue was taken up in H_2O and K_2CO_3 (3.46 g, 25.00 mmol, 2.5 eq) was added. A solution of Alloc-Cl (1.45 g, 12.00 mmol, 1.2 eq) in Et_2O was added to the flask and the biphasic mixture was stirred at rt overnight. The phases were separated and the aqueous phase was washed with EtOAc (3 x 10 mL) and the organic phases were discarded. The pH was adjusted to 1 and the resulting solution was extracted with EtOAc (3 x 15 mL). The combined organic phases were washed with brine and dried over $MgSO_4$. The product was obtained as a colorless oil (2.90 g, 9.16 mmol, 92%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.44 (s, 9H), 1.52 – 1.77 (m, 3H), 1.81 – 2.01 (m, 1H), 3.13 – 3.35 (m, 2H), 4.28 – 4.46 (m, 1H), 4.50 – 4.69 (m, 2H), 4.94 – 5.12 (m, 1H), 5.13 – 5.43 (m, 3H), 5.81 – 6.00 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 26.0, 28.4, 29.8, 40.5, 53.1, 65.8, 80.4, 117.9, 132.9, 155.9, 156.8, 175.9. **MS (ESI-)** *m/z* calculated for $C_{14}H_{23}N_2O_6$ $[M-H]^-$: 315.35; found 315.25.

*Boc-Lys(Alloc)-OH (***S5***):* Boc-Lys(Cbz)-OH (2.66 g, 7.00 mmol, 1.0 eq) and ammonium formate (2.21 g, 35.00 mmol, 5.0 eq) were dissolved in MeOH. Palladium on activated charcoal (10% Pd, 745 mg, 0.70 mmol, 10 mol%) was added and the resulting suspension was stirred for 2 h at rt. The mixture was filtered over Celite and the solvent was removed under reduced pressure. The residue was taken up in H_2O and K_2CO_3 (2.42 g, 17.50 mmol, 2.5 eq) was added. A solution of Alloc-Cl (1.01 g, 8.40 mmol, 1.2 eq) in Et_2O was added to the flask and the biphasic mixture was stirred at rt overnight. The phases were separated, the aqueous phase was washed with EtOAc (3 x 10 mL) and the organic phases were discarded. The pH was adjusted to 1 and the resulting solution was extracted with EtOAc (3 x 15 mL). The combined organic phases were washed with brine and dried over $MgSO_4$. The product was obtained as a colorless oil (1.36 g, 4.11 mmol, 59%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.34 – 1.61 (m, 13H), 1.66 – 1.92

(m, 2H), 3.18 (s, 2H), 4.30 (s, 1H), 4.51 – 4.65 (m, 2H), 4.90 – 5.14 (m, 1H), 5.15 – 5.36 (m, 2H), 5.83 – 6.01 (m, 1H), 6.31 (bs, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.4, 28.5, 29.5, 32.0, 40.6, 53.3, 65.8, 80.3, 117.9, 133.0, 156.0, 156.8, 176.4. **MS (ESI+)** *m/z* calculated for $C_{15}H_{30}N_3O_6$ $[M+NH_4]^+$: 348.21; found 348.26.

*Boc-Orn(Alloc)-OAll (***S6***):* Synthesized after general procedure 1.1. Boc-Orn(Alloc)-OH (**S7**, 2.53 g, 8.00 mmol), Cs_2CO_3 (2.61 g, 8.00 mmol), allyl bromide (1.26 g, 10.40 mmol). The product was obtained as a yellow oil (2.18 g, 6.12 mmol, 77%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.44 (s, 9H), 1.51 – 1.73 (m, 3H), 1.76 – 1.93 (m, 1H), 3.21 (dd, *J* = 12.7, 6.4 Hz, 2H), 4.26 – 4.40 (m, 1H), 4.50 – 4.69 (m, 4H), 4.81 (s, 1H), 5.07 (d, *J* = 8.4 Hz, 1H), 5.17 – 5.40 (m, 4H), 5.83 – 6.06 (m, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 26.1, 28.4, 30.2, 40.6, 53.2, 65.6, 66.1, 80.1, 117.8, 119.1, 131.6, 133.0, 155.5, 156.4, 172.4. **MS (ESI+)** *m/z* calculated for $C_{15}H_{30}N_3O_6$ $[M+NH_4]^+$: 374.23; found 374.23.

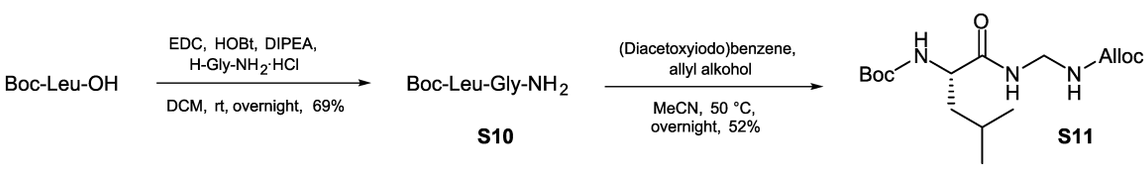
*Boc-Lys(Alloc)-OAll (***S7***):* Synthesized after general procedure 1.1. Boc-Lys(Alloc)-OH (**S8**, 1.32 g, 4.00 mmol), Cs_2CO_3 (1.30 g, 4.00 mmol), allyl bromide (627 g, 5.20 mmol). The product was obtained as a yellow oil (1.13 g, 3.06 mmol, 76%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.34 – 1.47 (m, 11H), 1.48 – 1.58 (m, 2H), 1.61 – 1.72 (m, 1H), 1.77 – 1.90 (m, 1H), 3.13 – 3.23 (m, 2H), 4.24 – 4.37 (m, 1H), 4.52 – 4.68 (m, 4H), 4.78 (bs, 1H), 5.08 (d, *J* = 6.9 Hz, 1H), 5.17 – 5.39 (m, 4H), 5.85 – 5.99 (m, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.5, 28.4, 29.5, 32.5, 40.7, 53.4, 65.6, 66.0, 80.1, 117.8, 119.0, 131.7, 133.1, 155.6, 156.5, 172.6. **MS (ESI+)** *m/z* calculated for $C_{18}H_{34}N_3O_6$ $[M+NH_4]^+$: 388.24; found 388.25.

Boc-Leu-Orn(Alloc)-OAll (**S8**): Synthesized after general procedure 1.2. Boc-Orn(Alloc)-OAll (**S9**, 713 mg, 2.00 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Leu-OH (463 mg, 2.00 mmol), EDC (498 mg, 2.60 mmol), HOBt (351 mg, 2.60 mmol) and DIPEA (646 mg, 5.00 mmol). The product was obtained as a yellow oil (925 mg, 1.97 mmol, 98%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.91 – 0.97 (m, 6H), 1.43 (s, 9H), 1.46 – 1.59 (m, 2H), 1.61 – 1.79 (m, 4H), 1.87 – 1.98 (m, 1H), 3.16 – 3.25 (m, 2H), 4.07 – 4.16 (m, 1H), 4.51 – 4.67 (m, 5H), 4.81 – 5.02 (m, 2H), 5.16 – 5.38 (m, 4H), 5.82 – 5.97 (m, 2H), 6.75 (d, *J* = 8.0 Hz, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 23.0, 23.2, 24.8, 25.9, 28.4, 29.6, 40.4, 41.1, 52.0, 53.3, 65.7, 66.2, 80.3, 117.8, 119.3, 131.5, 133.1, 155.9, 156.5, 171.7, 172.7. **MS (ESI+)** *m/z* calculated for $C_{23}H_{40}N_3O_7$ $[M+H]^+$: 470.59; found 470.43.

Boc-Leu-Lys(Alloc)-OAll (**S9**): Synthesized after general procedure 1.2. Boc-Lys(Alloc)-OAll (**S10**, 741 mg, 2.00 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Leu-OH (463 mg, 2.00 mmol), EDC (498 mg, 2.60 mmol), HOBt (351 mg, 2.60 mmol) and DIPEA (646 mg, 5.00 mmol). The product was obtained as a yellow oil (796 mg, 1.65 mmol, 82%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.90 – 0.98 (m, 6H), 1.30 – 1.40 (m, 2H), 1.43 (s, 9H), 1.46 – 1.58 (m, 2H), 1.61 – 1.76 (m, 4H), 1.82 – 1.93 (m, 1H), 3.09 – 3.24 (m, 2H), 4.10 – 4.20 (m, 1H), 4.53 – 4.69 (m, 5H), 4.98 – 5.15 (m, 2H), 5.18 – 5.38 (m, 4H), 5.83 – 5.98 (m, 2H), 6.69 (d, *J* = 7.1 Hz, 1H).

^{13}C NMR (101 MHz, $CDCl_3$) δ = 22.0, 22.4, 23.1, 24.8, 28.4, 29.1, 32.0, 40.5, 41.2, 52.0, 53.1, 65.6, 66.1, 80.2, 117.9, 119.2, 131.6, 133.1, 156.0, 156.5, 171.9, 173.0. **MS (ESI+)** *m/z* calculated for $C_{24}H_{42}N_3O_7$ $[M+H]^+$: 484.61; found 484.52.

Synthesis of Peptide S11

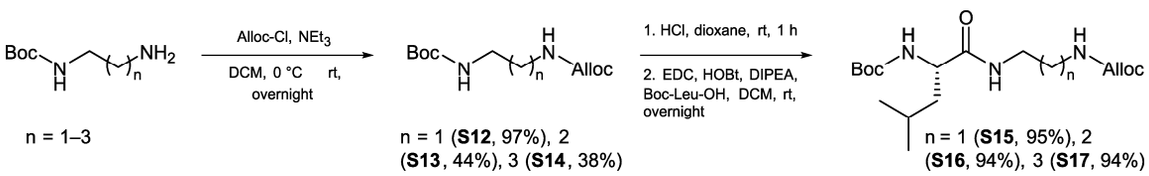


Scheme S2.3. Synthesis of peptide **S11**.

*Boc-Leu-Gly-NH_2 (**S10**):* Synthesized after general procedure 1.2. Boc-Leu-OH (2.31 g, 10.00 mmol, 1.0 eq), glycinamide hydrochloride (1.11 g, 10.00 mmol, 1.0 eq), EDC (2.49 g, 13.00 mmol, 1.3 eq), HOBt (1.76 g, 13.00 mmol, 1.3 eq) and DIPEA (3.23 g, 25.00 mmol, 2.5 eq) were dissolved in DCM and the mixture was stirred at rt overnight. The solution was diluted with EtOAc and extracted with 1 M HCl (3x15 mL) and sat. $NaHCO_3$ (3x15 mL). The combined organic phases were washed with brine and dried over $MgSO_4$. The crude product was purified by silica gel column chromatography (DCM/MeOH 50:1). The product was obtained as a colorless foam (2.00 g, 6.94 mmol, 69%). **^{1}H NMR** (400 MHz, DMSO-D6) δ = 0.83 – 0.89 (m, 6H), 1.38 (s, 9H), 1.40 – 1.46 (m, 2H), 1.54 – 1.66 (m, 1H), 3.52 – 3.70 (m, 2H), 3.91 (dd, *J* = 14.7, 7.4 Hz, 1H), 7.01 (d, *J* = 7.7 Hz, 1H), 7.07 (s, 1H), 7.14 (s, 1H), 7.98 – 8.04 (m, 1H). **^{13}C NMR** (101 MHz, DMSO-D6) δ = 21.6, 23.0, 24.2, 28.2, 40.3, 42.0, 53.0, 78.2, 155.7, 170.9, 172.8.

tert-butyl (S)-(1-(((((allyloxy)carbonyl)amino)methyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**S11**): Boc-Leu-Gly-NH_2 (**S13**, 287 mg, 1.00 mmol, 1.0 eq) and (Diacetoxyiodo)benzene (419 mg, 1.30 mmol, 1.3 eq) were dissolved in a mixture of MeCN (5 mL) and allyl alcohol (5 mL) and heated to 50 °C overnight. After removal of the solvent the crude product was purified by silica gel chromatography (DCM/MeOH 50:1). The product was obtained as a colorless solid (178 mg, 0.52 mmol, 52%); mp 151–152 °C. **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.91 – 0.96 (m, 6H), 1.43 (s, 9H), 1.58 – 1.75 (m, 3H), 4.10 (bs, 1H), 4.53 – 4.61 (m, 4H), 4.89 (d, *J* = 8.1 Hz, 1H), 5.19 – 5.34 (m, 2H), 5.79 – 5.95 (m, 2H), 7.13 (bs, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.0, 23.1, 24.8, 28.4, 41.5, 46.4, 53.1, 65.9, 80.2, 118.0, 132.6, 155.7, 156.7, 174.1. **MS (ESI+)** *m/z* calculated for $C_{16}H_{33}N_4O_5$ $[M+NH_4]^+$: 361.25; found: 361.26.

Synthesis of Peptides S15–S17



Scheme S2.4. Synthesis of peptides **S15–S17**.

*allyl tert-butyl ethane-1,2-diyldicarbamate (***S12***):* Synthesized after general procedure 1.3. *tert*-Butyl (2-aminoethyl)carbamate (481 mg, 3.00 mmol), NEt_3 (607 mg, 6.00 mmol), Alloc-Cl (434 mg, 3.60 mmol). The product was obtained as a colorless solid (708 mg, 2.90 mmol, 97%); mp 107–108 °C. **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.43 (s, 9H), 3.16 – 3.38 (m, 4H), 4.55 (d, *J* = 5.5 Hz, 2H), 4.89 (s, 1H), 5.13 – 5.23 (m, 2H), 5.29 (dd, *J* = 17.2, 1.4 Hz, 1H), 5.82 – 5.98 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 28.5, 40.7, 41.6, 65.7, 79.7, 117.8, 132.9, 156.5, 156.8. **MS (ESI+)** *m/z* calculated for $C_{11}H_{24}N_3O_4$ $[M+H]^+$: 262.18; found 262.22.

*allyl tert-butyl propane-1,3-diyldicarbamate (***S13***):* Synthesized after general procedure 1.3. *tert*-Butyl (2-aminopropyl)carbamate (523 mg, 3.00 mmol), NEt_3 (607 mg, 6.00 mmol), Alloc-Cl (434 mg, 3.60 mmol). The product was obtained as a colorless oil (342 mg, 1.32 mmol, 44%). **^{1}H NMR** (400 MHz, DMSO-D6) δ = 1.37 (s, 9H), 1.44 – 1.55 (m, 2H), 2.86 – 3.00 (m, 4H), 4.45 (d, *J* = 5.3 Hz, 2H), 5.12 – 5.19 (m, 1H), 5.22 – 5.31 (m, 1H), 5.84 – 5.98 (m, 1H), 6.74 (bs, 1H), 7.12 (bs, 1H). **^{13}C NMR** (101 MHz, DMSO-D6) δ = 28.2, 29.9, 37.5, 38.0, 64.1, 77.4, 116.8, 133.8, 155.6, 155.9. **MS (ESI+)** *m/z* calculated for $C_{12}H_{23}N_2O_4$ $[M+H]^+$: 259.33; found 259.22.

*allyl tert-butyl butane-1,3-diyldicarbamate (***S14***):* Synthesized after general procedure 1.3. *tert*-Butyl (2-aminobutyl)carbamate (565 mg, 3.00 mmol), NEt_3 (607 mg, 6.00 mmol), Alloc-Cl (434 mg, 3.60 mmol). The product was obtained as a colorless solid (310 mg, 1.14 mmol, 38%); mp 100–101 °C. **^{1}H NMR** (400 MHz, DMSO-D6) δ = 1.29 – 1.43 (m, 15H), 2.84 – 3.09 (m, 4H), 4.45 (d, *J* = 5.2 Hz, 2H), 5.16 (dd, *J* = 10.6, 0.9 Hz, 1H), 5.21 – 5.34 (m, 1H), 5.82 – 6.03 (m, 1H), 6.70 – 6.87 (m, 1H), 7.07 – 7.23 (m, 1H). **^{13}C NMR** (101 MHz, DMSO-D6) δ = 26.8, 28.3, 64.1, 77.3, 116.8, 133.9, 155.6, 155.9. **MS (ESI+)** *m/z* calculated for $C_{13}H_{25}N_2O_4$ $[M+H]^+$: 273.35; found 273.22.

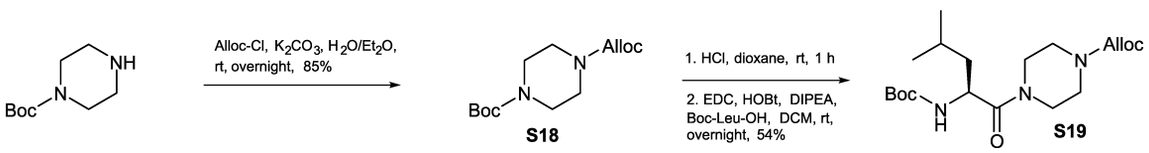
*tert-butyl (S)-(1-((2-(((allyloxy)carbonyl)amino)ethyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (***S15***):* Synthesized after general procedure 1.2. **S12** (489 mg, 2.00 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Leu-OH (463 mg, 2.00 mmol), EDC (498 mg, 2.60 mmol, 1.3 eq), HOBt (351 mg, 2.60 mmol) and DIPEA (569 mg, 4.4 mmol). The product

was obtained as a yellow oil (681 mg, 1.91 mmol, 95%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.88 – 0.95 (m, 6H), 1.43 (s, 9H), 1.60 – 1.78 (m, 3H), 3.27 – 3.48 (m, 4H), 4.04 (dd, *J* = 12.9, 8.1 Hz, 1H), 4.55 (d, *J* = 5.5 Hz, 2H), 4.91 (d, *J* = 6.7 Hz, 1H), 5.17 – 5.38 (m, 3H), 5.84 – 5.98 (m, 1H), 6.60 (bs, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.1, 23.1, 24.9, 28.4, 40.1, 40.9, 41.3, 53.5, 65.8, 80.4, 117.8, 132.9, 156.0, 157.0, 173.6. **MS (ESI+)** *m/z* calculated for $C_{17}H_{32}N_3O_5$ $[M+H]^+$: 358.46; found 358.23.

*tert-butyl (S)-(1-((2-(((allyloxy)carbonyl)amino)propyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (***S16***):* Synthesized after general procedure 1.2. **S13** (300 mg, 1.16 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Leu-OH (269 mg, 1.16 mmol), EDC (284 mg, 1.48 mmol), HOBt (200 mg, 1.48 mmol) and DIPEA (442 mg, 3.42 mmol). The product was obtained as a colorless oil (407 mg, 1.10 mmol, 94%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.80 – 1.06 (m, 6H), 1.31 – 1.52 (m, 9H), 1.55 – 1.73 (m, 4H), 1.82 – 2.02 (m, 1H), 3.11 – 3.42 (m, 4H), 4.05 (bs, 1H), 4.55 (d, *J* = 5.5 Hz, 2H), 4.92 (s, 1H), 5.10 – 5.50 (m, 3H), 5.78 – 6.00 (m, 1H), 6.64 (bs, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.2, 23.1, 24.9, 28.4, 30.1, 36.0, 37.5, 41.4, 53.4, 65.6, 80.3, 117.7, 133.1, 155.9, 156.9, 173.4. **MS (ESI+)** *m/z* calculated for $C_{18}H_{34}N_3O_5$ $[M+H]^+$: 372.49; found 372.41.

*tert-butyl (S)-(1-((2-(((allyloxy)carbonyl)amino)butyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (***S17***):* Synthesized after general procedure 1.2. **S14** (272 mg, 1.00 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Leu-OH (231 mg, 1.00 mmol), EDC (249 mg, 1.30 mmol), HOBt (176 mg, 1.30 mmol) and DIPEA (388 mg, 3.00 mmol). The product was obtained as a colorless solid (364 mg, 0.94 mmol, 94%); mp 80–81 °C. **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.85 – 1.01 (m, 6H), 1.42 (s, 9H), 1.49 – 1.56 (m, 4H), 1.58 – 1.70 (m, 2H), 1.83 – 2.01 (m, 1H), 3.10 – 3.38 (m, 4H), 4.06 (bs, 1H), 4.55 (d, *J* = 5.3 Hz, 2H), 4.96 (bs, 2H), 5.19 (dd, *J* = 10.4, 1.2 Hz, 1H), 5.25 – 5.34 (m, 1H), 5.84 – 6.05 (m, 1H), 6.42 (bs, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.1, 23.0, 24.8, 26.7, 27.3, 28.4, 39.1, 40.6, 41.3, 53.2, 65.6, 80.1, 117.7, 133.0, 155.9, 156.5, 172.8. **MS (ESI+)** *m/z* calculated for $C_{19}H_{36}N_3O_5$ $[M+H]^+$: 386.51; found 386.36.

Synthesis of Peptide S19

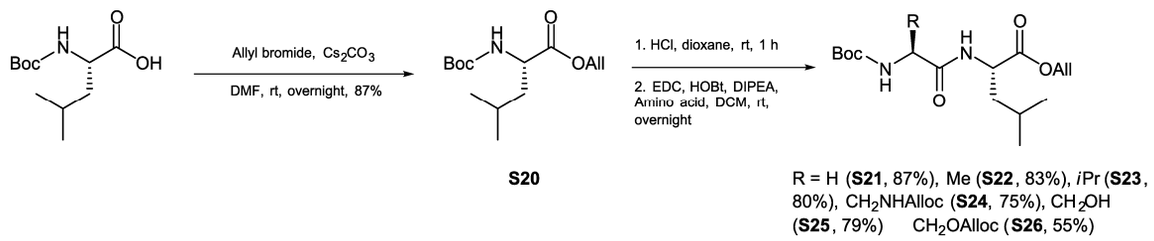


Scheme S2.5. Synthesis of peptide **S19**.

Piperazine **S18**: Boc-Piperazine (1.12 g, 6.00 mmol, 1.0 eq) and NEt_3 (729 mg, 7.20 mmol, 1.2 eq) were dissolved in DCM. The solution was cooled to 0 °C and Alloc-Cl (868 mg, 7.20 mmol, 1.2 eq) was added dropwise. The mixture was allowed to warm to rt overnight and was washed with 1 M HCl (3 x 10 mL), brine and dried over anhydrous $MgSO_4$. The product was obtained as a yellow oil (1.38 g, 5.10 mmol, 85%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.46 (s, 9H), 3.15 – 3.64 (m, 8H), 4.46 – 4.75 (m, 2H), 5.17 – 5.33 (m, 2H), 5.73 – 6.10 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 28.5, 43.5, 66.2, 80.3, 117.8, 133.0, 154.5, 155.2. **MS (ESI+)** *m/z* calculated for $C_{13}H_{26}N_3O_4$ $[M+NH_4]^+$: 288.37; found 288.20.

Amide **S19**: Synthesized after general procedure 1.2. **S18** (541 mg, 2.00 mmol), HCl (4 M solution in dioxane, 2 mL, 8.00 mmol), Boc-Leu-OH (463 mg, 2.00 mmol), EDC (498 mg, 2.60 mmol), HOBt (352 mg, 2.60 mmol) and DIPEA (810 mg, 8.00 mmol). The product was obtained as a yellow oil (415 mg, 1.08 mmol, 54%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.85 – 1.06 (m, 6H), 1.43 (s, 9H), 1.45 – 1.55 (m, 2H), 1.64 – 1.76 (m, 1H), 3.37 – 3.82 (m, 7H), 4.53 – 4.70 (m, 2H), 5.13 – 5.41 (m, 2H), 5.85 – 6.03 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.1, 23.5, 24.8, 28.5, 42.0, 43.0, 45.4, 48.5, 49.7, 66.6, 79.9, 118.1, 132.8, 155.1, 155.7, 171.8. **MS (ESI+)** *m/z* calculated for $C_{19}H_{34}N_3O_5$ $[M+H]^+$: 384.50; found 384.34.

Synthesis of Peptides S21–S26



Scheme S2.6. Synthesis of peptides **S21–S26**.

Boc-Leu-OAll (**S20**): Synthesized after general procedure 1.1. Boc-Leu-OH (5.00 g, 21.62 mmol, 1.0 eq) and Cs_2CO_3 (7.04 g, 21.62 mmol, 1.0 eq) were dissolved in DMF. Allyl bromide (3.39 g, 28.11 mmol, 1.3 eq) was added and the solution was stirred for 5 h at rt. The solvent was removed under reduced pressure and the residual slurry taken up in EtOAc. The organic phase was washed with sat. $NaHCO_3$ solution (3 x 15 mL) and brine and dried over $MgSO_4$. The solvent was removed to give the product as a yellow oil (5.12 g, 18.85 mmol, 87%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.91 – 0.97 (m, 6H), 1.44 (s, 9H), 1.47 – 1.77 (m, 3H), 4.29 – 4.38 (m, 1H), 4.56 – 4.69 (m, 2H), 4.85 – 4.92 (m, 1H), 5.25 (dd, *J* = 10.4 Hz, 1.0, 1H), 5.29 – 5.39 (m, 1H), 5.83 – 5.98 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.0, 23.0, 24.9, 28.4, 42.0, 52.3, 65.9, 80.05, 118.7, 131.9, 155.6, 173.4. **MS (ESI+)** *m/z* calculated for $C_{14}H_{26}NO_4$ $[M+H]^+$: 272.36; found 272.16.

Boc-Gly-Leu-OAll (**S21**): Synthesized after general procedure 1.2. Boc-Leu-OAll (**S20**, 407 mg, 1.50 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Gly-OH (263 mg, 1.50 mmol), EDC (374 mg, 1.95 mmol), HOBt (264 mg, 1.95 mmol) and DIPEA (485 mg, 3.75 mmol). The product was obtained as a colorless oil (431 mg, 1.31 mmol, 87%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.88 – 0.97 (m, 6H), 1.45 (s, 9H), 1.50 – 1.74 (m, 3H), 3.73 – 3.94 (m, 2H), 4.59 – 4.73 (m, 3H), 5.08 – 5.19 (m, 1H), 5.21 – 5.40 (m, 2H), 5.81 – 5.98 (m, 1H), 6.44 – 6.58 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.0, 23.0, 24.9, 28.4, 41.6, 44.4, 50.8, 66.1, 80.4, 119.0, 131.7, 156.2, 169.5, 172.7. **MS (ESI+)** *m/z* calculated for $C_{16}H_{32}N_3O_5$ $[M+NH_4]^+$: 346.45; found 346.28.

Boc-Ala-Leu-OAll (**S22**): Synthesized after general procedure 1.2. Boc-Leu-OAll (**S20**, 407 mg, 1.50 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Ala-OH (284 mg, 1.50 mmol), EDC (374 mg, 1.95 mmol), HOBt (264 mg, 1.95 mmol) and DIPEA (485 mg, 3.75 mmol). The product was obtained as a colorless oil (426 mg, 1.28 mmol, 83%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.88 – 0.97 (m, 6H), 1.35 (d, *J* = 7.0 Hz, 3H), 1.44 (s, 9H), 1.52 – 1.76 (m,

3H), 4.12 – 4.23 (m, 1H), 4.55 – 4.69 (m, 3H), 4.99 (bs, 1H), 5.19 – 5.39 (m, 2H), 5.79 – 6.00 (m, 1H), 6.44 – 6.60 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 18.0, 21.9, 23.0, 24.9, 28.4, 41.6, 50.0, 50.8, 66.0, 80.2, 118.9, 131.7, 155.7, 172.6. **MS (ESI+)** *m/z* calculated for $C_{17}H_{34}N_3O_5$ $[M+NH_4]^+$: 360.47; found 360.34.

Boc-Val-Leu-OAll (**S23**): Synthesized after general procedure 1.2. Boc-Leu-OAll (**S20**, 407 mg, 1.50 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Val-OH (326 mg, 1.50 mmol), EDC (374 mg, 1.95 mmol), HOBt (264 mg, 1.95 mmol) and DIPEA (485 mg, 3.75 mmol). The product was obtained as a colorless solid (444 mg, 1.20 mmol, 80%); mp 93–94 °C. **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.88 – 1.01 (m, 12H), 1.44 (s, 9H), 1.53 – 1.73 (m, 3H), 2.04 – 2.20 (m, 1H), 3.85 – 3.97 (m, 1H), 4.54 – 4.71 (m, 3H), 4.98 – 5.12 (m, 1H), 5.21 – 5.42 (m, 2H), 5.81 – 5.98 (m, 1H), 6.12 – 6.32 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 18.0, 19.4, 21.9, 23.0, 24.9, 28.4, 31.0, 41.6, 50.8, 60.1, 66.0, 80.0, 119.0, 131.7, 156.0, 171.6, 172.5. **MS (ESI+)** *m/z* calculated for $C_{19}H_{36}N_3O_5$ $[M+NH_4]^+$: 388.53; found 388.38.

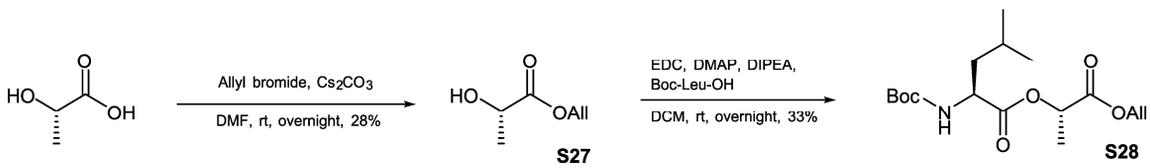
Boc-Dap(Alloc)-Leu-OAll (**S24**): Synthesized after general procedure 1.2. Boc-Leu-OAll (**S20**, 407 mg, 1.50 mmol), HCl (4 M solution in dioxane, 1 mL, 4.00 mmol), Boc-Dap(Alloc)-OH (432 mg, 1.50 mmol), EDC (374 mg, 1.95 mmol), HOBt (264 mg, 1.95 mmol) and DIPEA (485 mg, 3.75 mmol). The product was obtained as a colorless solid (494 mg, 1.12 mmol, 75%); mp 75–76 °C. **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.84 – 1.00 (m, 6H), 1.44 (s, 9H), 1.54 – 1.67 (m, 3H), 3.49 – 3.66 (m, 2H), 4.24 (bs, 1H), 4.51 – 4.69 (m, 5H), 5.18 – 5.40 (m, 4H), 5.51 (bs, 1H), 5.66 (bs, 1H), 5.83 – 6.01 (m, 2H), 7.04 (s, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 21.8, 23.0, 24.9, 28.4, 41.2, 42.9, 50.9, 51.1, 66.1, 80.7, 118.0, 119.0, 131.7, 132.7, 156.2, 157.5, 170.7, 172.3 **MS (ESI+)** *m/z* calculated for $C_{21}H_{39}N_4O_7$ $[M+NH_4]^+$: 459.56; found 459.40.

Boc-Ser-Leu-OAll (**S25**): Synthesized after general procedure 1.2. Boc-Leu-OAll (**S20**, 1.09 g, 4.00 mmol), HCl (4 M solution in dioxane, 2 mL, 8.00 mmol), Boc-Ser-OH (821 mg, 4.00 mmol), EDC (997 mg, 5.20 mmol), HOBt (703 mg, 5.20 mmol) and DIPEA (1.29 g, 10.00 mmol). The product was obtained as a colorless oil (1.14 g, 3.17 mmol, 79%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.87 – 1.01 (m, 6H), 1.45 (s, 9H), 1.53 – 1.74 (m, 3H), 3.63 (dd, *J* = 11.4, 5.5 Hz, 1H), 4.02 – 4.12 (m, 1H), 4.13 – 4.20 (m, 1H), 4.53 – 4.69 (m, 3H), 5.21 – 5.40 (m, 2H), 5.51 – 5.64 (m, 1H), 5.81 – 6.04 (m, 1H), 6.84 – 6.99 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 21.7, 23.0, 24.9, 28.4, 41.0, 51.2, 54.7, 63.1, 66.2, 80.6, 119.1, 131.6, 156.2, 171.7, 172.7. **MS (ESI+)** *m/z* calculated for $C_{17}H_{34}N_3O_6$ $[M+NH_4]^+$: 376.47; found 376.34.

Boc-Ser(OAlloc)-Leu-OAll (**S26**): Boc-Ser-Leu-OAll (**S25**, 358 mg, 1.00 mmol, 1.0 eq) was dissolved in dry DCM under argon and the solution was cooled to -50 °C. Pyridine (197 mg, 2.50 mmol, 2.5 eq) and Alloc-Cl (133 mg, 1.10 mmol, 1.1 eq) were added subsequently. The mixture was stirred at -50 °C for 2 h and allowed to warm to rt overnight. The solution was washed with 1 M HCl (3 x 10 mL), sat. $NaHCO_3$ solution (3 x 10 mL), brine and dried over anhydrous $MgSO_4$. The crude product was purified by silica gel chromatography

(DCM/MeOH 50:1) to give the pure product as a colorless oil (242 mg, 0.55 mmol, 55%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.89 – 0.97 (m, 6H), 1.46 (s, 9H), 1.54 – 1.71 (m, 3H), 4.31 (dd, *J* = 10.9, 4.8 Hz, 1H), 4.45 (s, 1H), 4.53 (dd, *J* = 10.9, 5.2 Hz, 1H), 4.59 – 4.71 (m, 5H), 5.20 – 5.42 (m, 4H), 5.83 – 5.99 (m, 2H), 6.68 (d, *J* = 7.2 Hz, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 25.8, 26.0, 28.4, 29.8, 40.5, 53.1, 54.3, 65.8, 66.1, 66.4, 80.4, 117.9, 118.2, 132.5, 132.9, 155.9, 156.8, 175.8, 175.9. **MS (ESI+)** *m/z* calculated for $C_{21}H_{38}N_3O_8$ $[M+NH_4]^+$: 460.55; found 460.30.

Synthesis of Peptide S28



Scheme S2.7. Synthesis of peptide **S28**.

*(S)-2-hydroxypropanoic acid allyl ester (***S27***)*: Synthesized after general procedure 1.1. L-lactic acid (720 mg, 8.00 mmol), Cs_2CO_3 (2.61 g, 8.00 mmol), allyl bromide (1.26 g, 10.40 mmol). The product was obtained as a yellow oil (296 mg, 2.27 mmol, 28%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.44 (d, *J* = 6.9 Hz, 3H), 4.30 (q, *J* = 6.9 Hz, 1H), 4.61 – 4.77 (m, 2H), 5.21 – 5.38 (m, 2H), 5.84 – 6.00 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 20.5, 66.2, 66.8, 119.0, 131.5, 175.5.

Ester **S28***:* Synthesized after general procedure 1.2. (S)-2-hydroxypropanoic acid allyl ester (**S27**, 260 mg, 2.00 mmol, 1.0 eq), Boc-Leu-OH (462 mg, 2.00 mmol, 1.0 eq), EDC (447 mg, 2.60 mmol, 1.3 eq) and DMAP (24 mg, 0.20 mmol, 0.1 eq) were suspended in DCM. DIPEA (646 mg, 5.00 mmol, 2.5 eq) was added and the resulting solution was stirred at rt overnight. The solution was washed with 1 M HCl (3 x 10 mL), sat. $NaHCO_3$ solution (3 x 10 mL), brine and dried over anhydrous $MgSO_4$. The crude product was purified by silica gel chromatography (DCM/MeOH 50:1) to give the pure product as a colorless oil (225 mg, 0.66 mmol, 33%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.89 – 1.00 (m, 6H), 1.43 (s, 9H), 1.48 – 1.57 (m, 4H), 1.69 – 1.89 (m, 2H), 4.31 – 4.44 (m, 1H), 4.62 (ddd, *J* = 5.8, 2.6, 1.2 Hz, 2H), 4.84 (d, *J* = 8.4 Hz, 1H), 5.15 (q, *J* = 7.1 Hz, 1H), 5.25 (ddd, *J* = 10.4, 2.5, 1.2 Hz, 1H), 5.32 (ddd, *J* = 17.2, 2.9, 1.5 Hz, 1H), 5.82 – 6.00 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 17.0, 21.7, 23.1, 24.8, 28.4, 41.6, 51.9, 66.0, 69.1, 80.0, 119.0, 131.5, 155.6, 170.3, 173.2. **MS (ESI+)** *m/z* calculated for $C_{17}H_{30}NO_6$ $[M+H]^+$: 344.43; found 344.26.

2.6.3 Determination of Purity by ^{1}H-qNMR

For the determination of purity by absolute ^{1}H-qNMR, an internal calibration with dimethyl terephthalate (DMT) was employed. The spectra were recorded in deuterated methanol with number of scans set to 32 and a relaxation delay of 50–60 s. Purity was calculated as follows [106]:

$$\boldsymbol{P}[\%] = \frac{\boldsymbol{n_{IC}} \cdot \boldsymbol{I_a} \cdot \boldsymbol{M_a} \cdot \boldsymbol{m_{IC}} \cdot \boldsymbol{P_{IC}}}{\boldsymbol{n_a} \cdot \boldsymbol{I_{IC}} \cdot \boldsymbol{M_{IC}} \cdot \boldsymbol{m_a}}$$

Were n is the number of protons, I is the integral, M is the molecular weight, m is the mass and P is the purity (in %). Indices IC and a denominate internal calibrant or analyte respectively. For the determination of analyte integral, the sum of suitable, non-overlapping signals was used. The OCH_3 signal of DMT was not considered for quantitative analysis, since transesterification with deuterated methanol can result in artificial signal reduction.

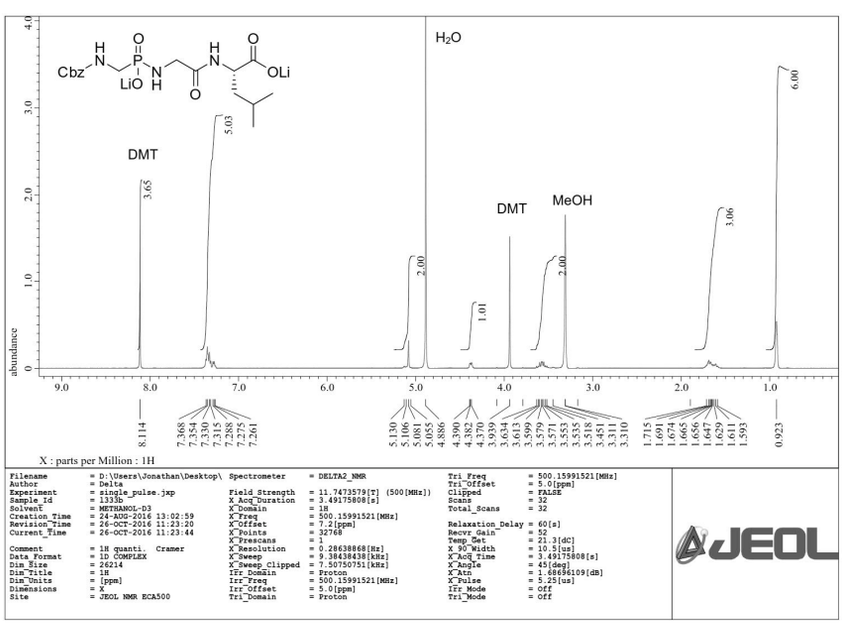


Figure S2.1. ^{1}H-qNMR spectrum of **26**.

$$\boldsymbol{m_a = 3.006\,mg;\ M_a = 427.24\,g/mol;\ I_a = 19.10;\ n_a = 19}$$

$$\boldsymbol{m_{IC} = 1.206\,mg;\ M_{IC} = 194.18\,g/mol;\ I_{IC} = 3.65;\ n_{IC} = 4}$$

$$\boldsymbol{P[\%] = \frac{n_{IC} \cdot I_a \cdot M_a \cdot m_{IC} \cdot P_{IC}}{n_a \cdot I_{IC} \cdot M_{IC} \cdot m_a}}$$

$$\boldsymbol{P[\%] = \frac{4 \cdot 19.10 \cdot 427.24\,g \cdot mol^{-1} \cdot 1.206\,mg \cdot 99.99\%}{19 \cdot 3.65 \cdot 194.18\,g \cdot mol^{-1} \cdot 3.006\,mg} = 97.24\%}$$

2.6.4 NMR Spectra of 2, 3, 18–31

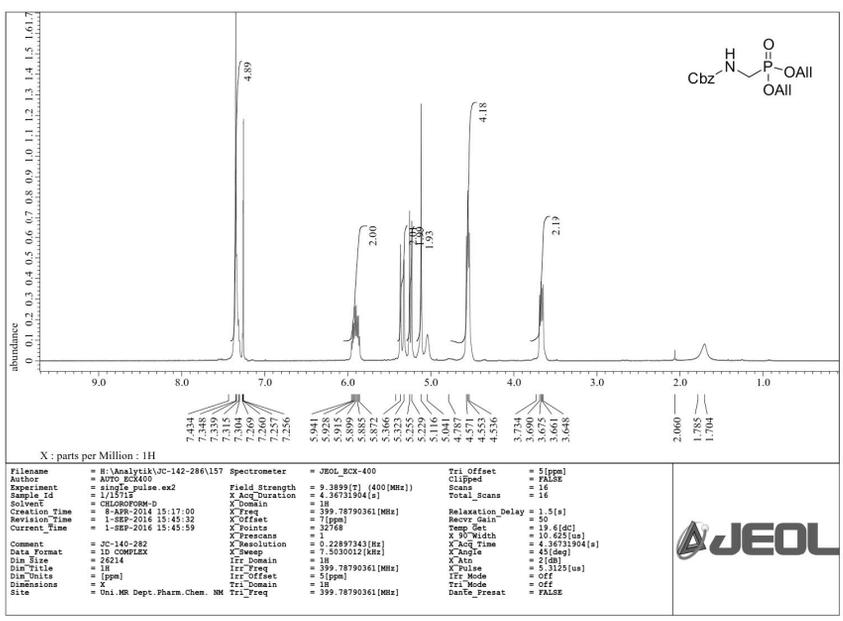


Figure S2.2. ^{1}H NMR Spectrum of **2**.

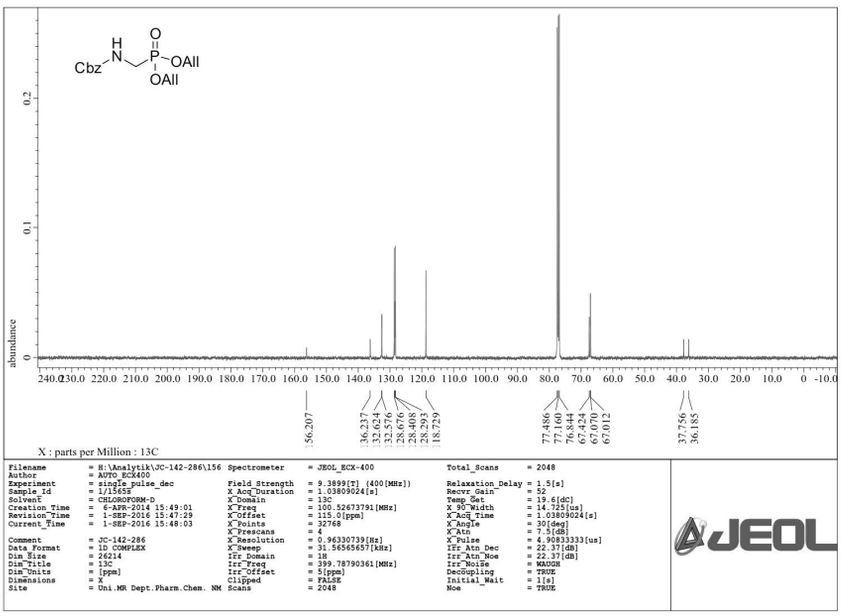


Figure S2.3. ^{13}C NMR Spectrum of **2**.

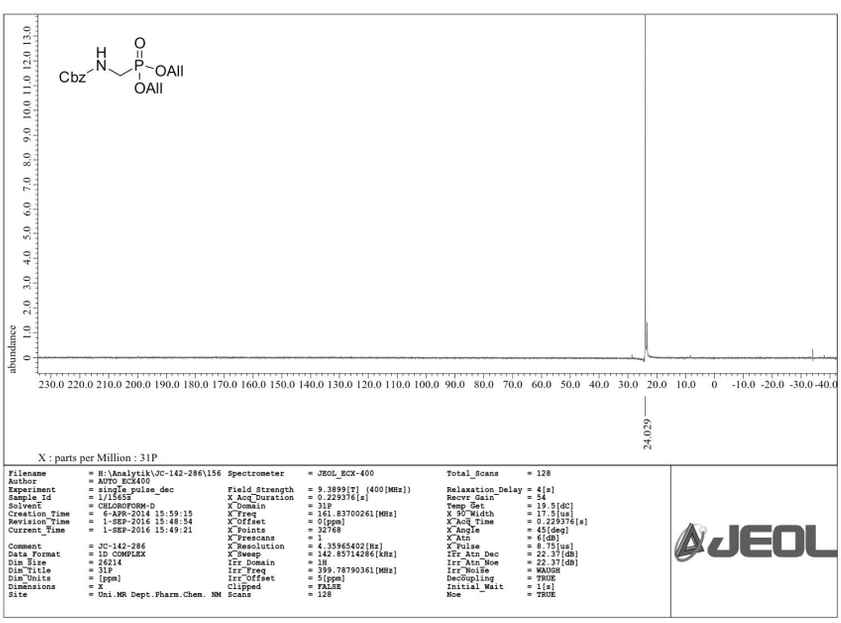


Figure S2.4. ^{31}P NMR Spectrum of **2**.

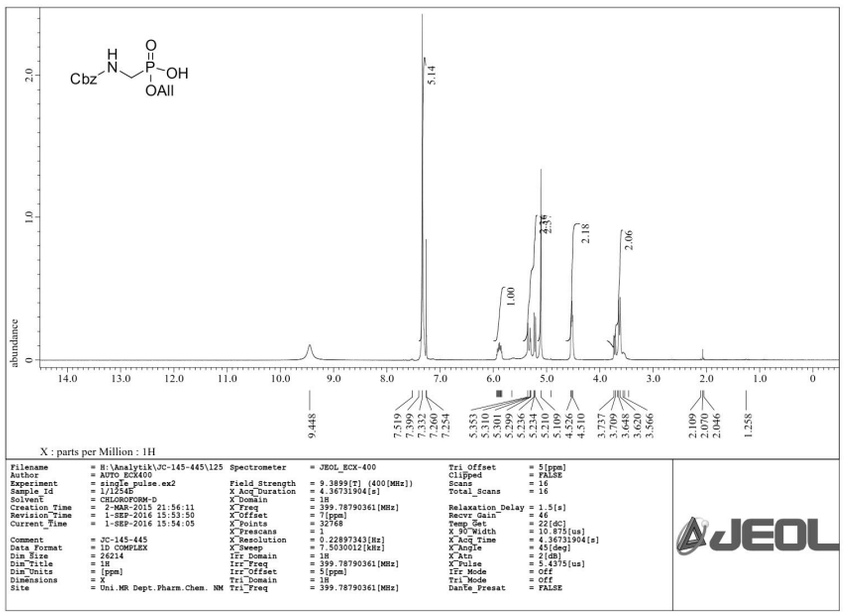


Figure S2.5. [1]H NMR Spectrum of **3**.

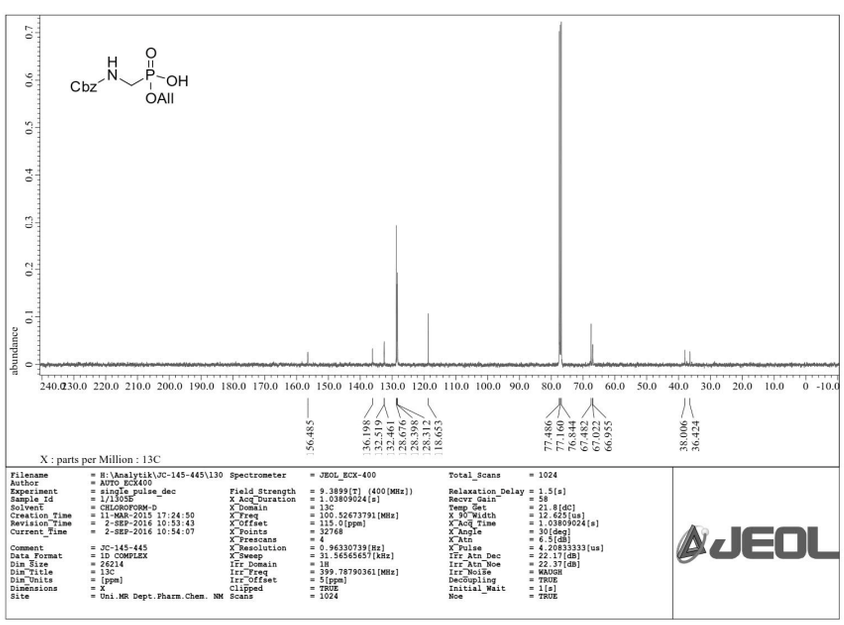


Figure S2.6. [13]C NMR Spectrum of **3**.

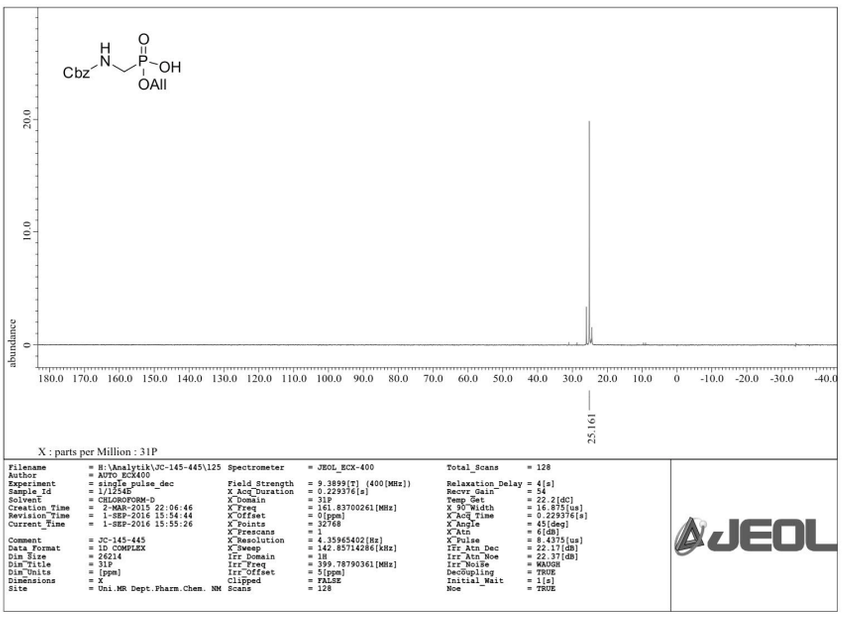


Figure S2.7. ^{31}P NMR Spectrum of **3**.

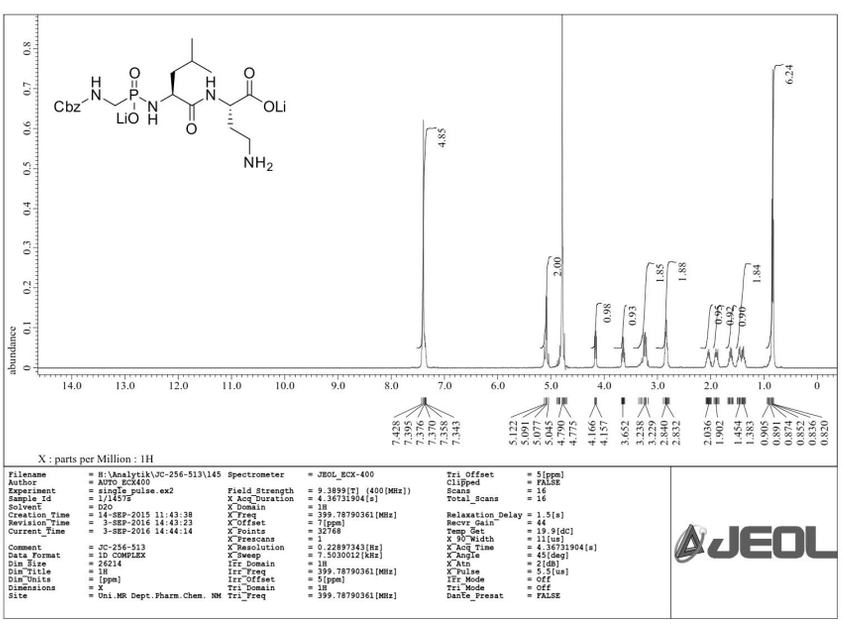


Figure S2.8. ^{1}H NMR Spectrum of **18**.

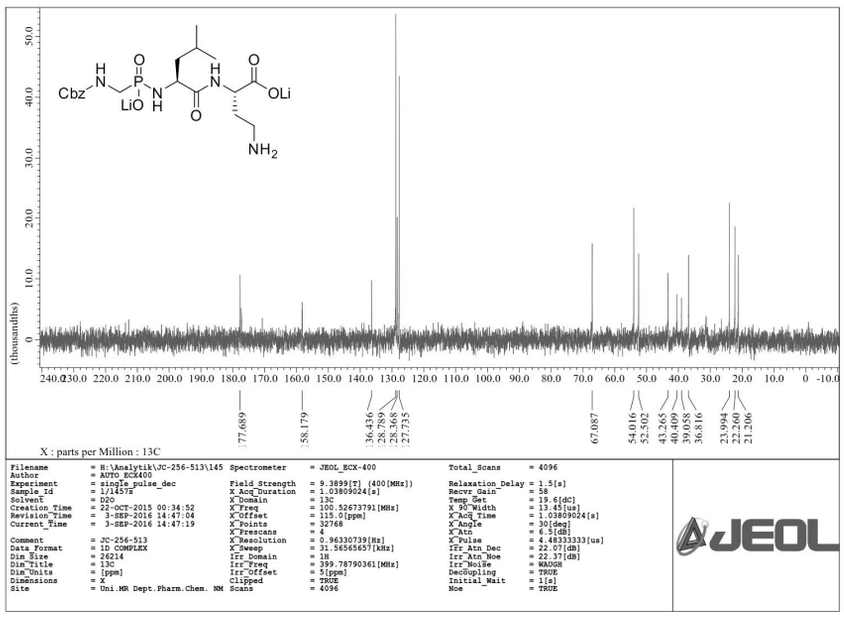


Figure S2.9. ^{13}C NMR Spectrum of **18**.

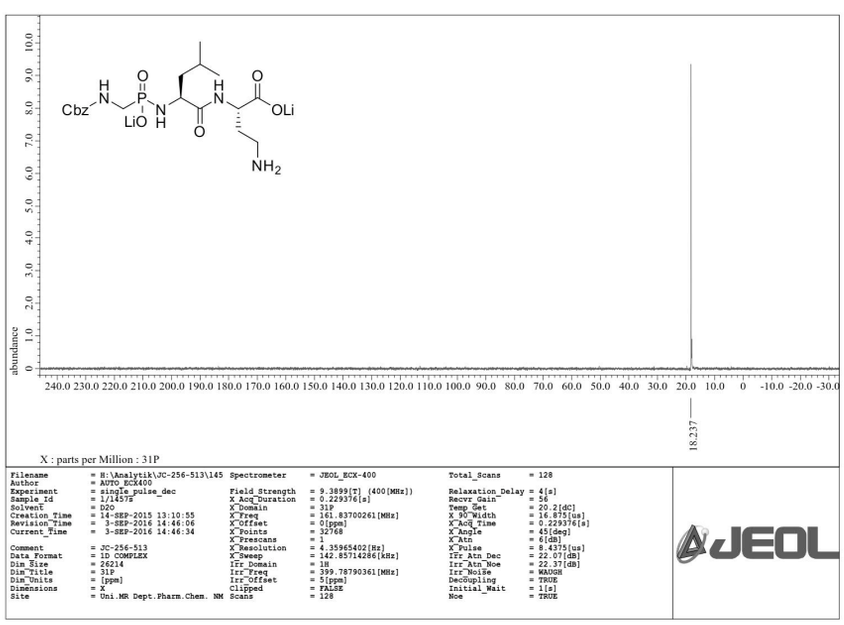
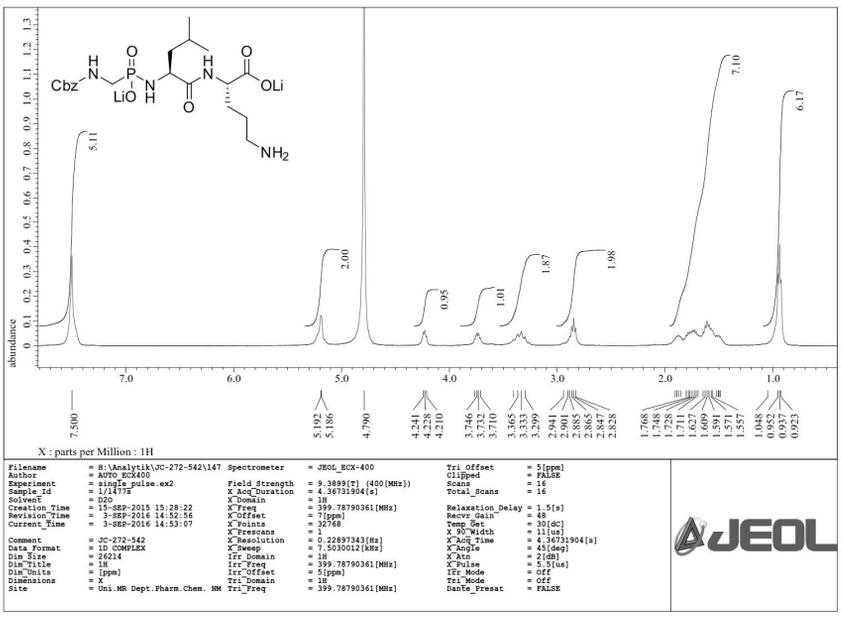
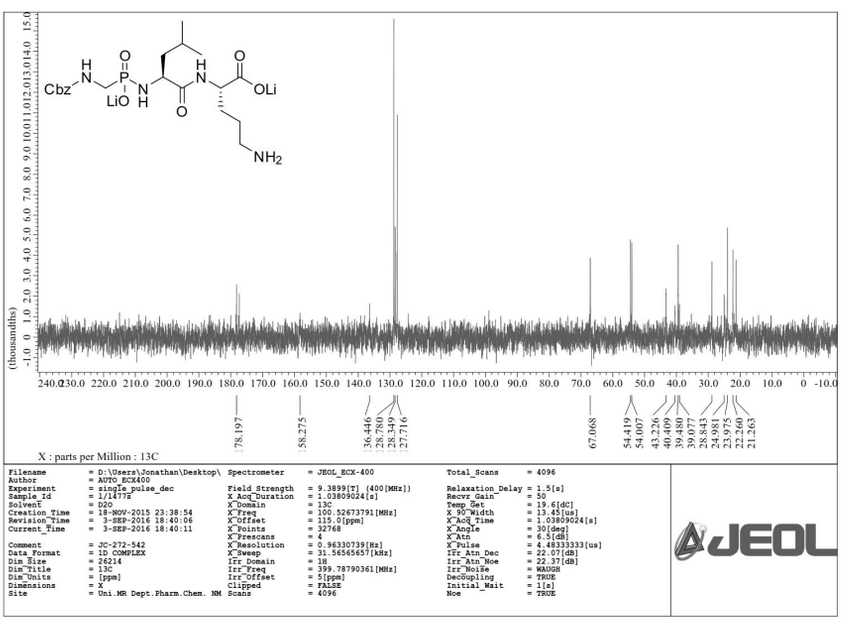


Figure S2.10. ^{31}P NMR Spectrum of **18**.



Filename	= H:\Analytik\JC-272-542\147	Spectrometer	= JEOL_ECX-400	Tri_Offset	= 5[ppm]
Author	= AUTO ECX400			Clipped	= FALSE
Experiment	= single_pulse.ex2	Field_Strength	= 9.3899[T] (400[MHz])	Scans	= 16
Sample_Id	= 1/1477s	X_Acq_Duration	= 4.36731904[s]	Total_Scans	= 16
Solvent	= D2O	X_Domain	= 1H		
Creation_Time	= 15-SEP-2015 15:28:22	X_Freq	= 399.78790361[MHz]	Relaxation_Delay	= 1.5[s]
Revision_Time	= 3-SEP-2016 14:52:56	X_Offset	= 7[ppm]	Recvr_Gain	= 48
Current_Time	= 3-SEP-2016 14:53:07	X_Points	= 32768	Temp_Get	= 30[dC]
		X_Prescans	= 1	X_90_Width	= 11[us]
Comment	= JC-272-542	X_Resolution	= 0.22897343[Hz]	X_Acq_Time	= 4.36731904[s]
Data_Format	= 1D COMPLEX	X_Sweep	= 7.5030012[kHz]	X_Angle	= 45[deg]
Dim_Size	= 26214	IFr_Domain	= 1H	X_Atn	= 2[dB]
Dim_Title	= 1H	Irr_Freq	= 399.78790361[MHz]	X_Pulse	= 5.5[us]
Dim_Units	= [ppm]	Irr_Offset	= 5[ppm]	IFr_Mode	= Off
Dimensions	= X	Tri_Domain	= 1H	Tri_Mode	= Off
Site	= Uni.MR Dept.Pharm.Chem. NM	Tri_Freq	= 399.78790361[MHz]	Dante_Presat	= FALSE

Figure S2.11. ^{1}H NMR Spectrum of **19**.



Filename	= D:\Users\Jonathan\Desktop\	Spectrometer	= JEOL_ECX-400	Total_Scans	= 4096
Author	= AUTO_ECX400				
Experiment	= single_pulse_dec	Field_Strength	= 9.3899[T] (400[MHz])	Relaxation_Delay	= 1.5[s]
Sample_Id	= 1/1477s	X_Acq_Duration	= 1.03809024[s]	Recvr_Gain	= 50
Solvent	= D2O	X_Domain	= 13C	Temp_Get	= 19.6[dC]
Creation_Time	= 18-NOV-2015 23:38:54	X_Freq	= 100.52673791[MHz]	X_90_Width	= 13.45[us]
Revision_Time	= 3-SEP-2016 18:40:06	X_Offset	= 115.0[ppm]	X_Acq_Time	= 1.03809024[s]
Current_Time	= 3-SEP-2016 18:40:11	X_Points	= 32768	X_Angle	= 30[deg]
		X_Prescans	= 4	X_Atn	= 6.5[dB]
Comment	= JC-272-542	X_Resolution	= 0.96330739[Hz]	X_Pulse	= 4.48333333[us]
Data_Format	= 1D COMPLEX	X_Sweep	= 31.56565657[kHz]	IFr_Atn_Dec	= 22.07[dB]
Dim_Size	= 26214	IFr_Domain	= 1H	Irr_Atn_Noe	= 22.37[dB]
Dim_Title	= 13C	Irr_Freq	= 399.78790361[MHz]	Irr_NoiSe	= WALTZ16
Dim_Units	= [ppm]	Irr_Offset	= 5[ppm]	DecOupling	= TRUE
Dimensions	= X	Clipped	= FALSE	Initial_Wait	= 1[s]
Site	= Uni.MR Dept.Pharm.Chem. NM	Scans	= 4096	Noe	= TRUE

Figure S2.12. ^{13}C NMR Spectrum of **19**.

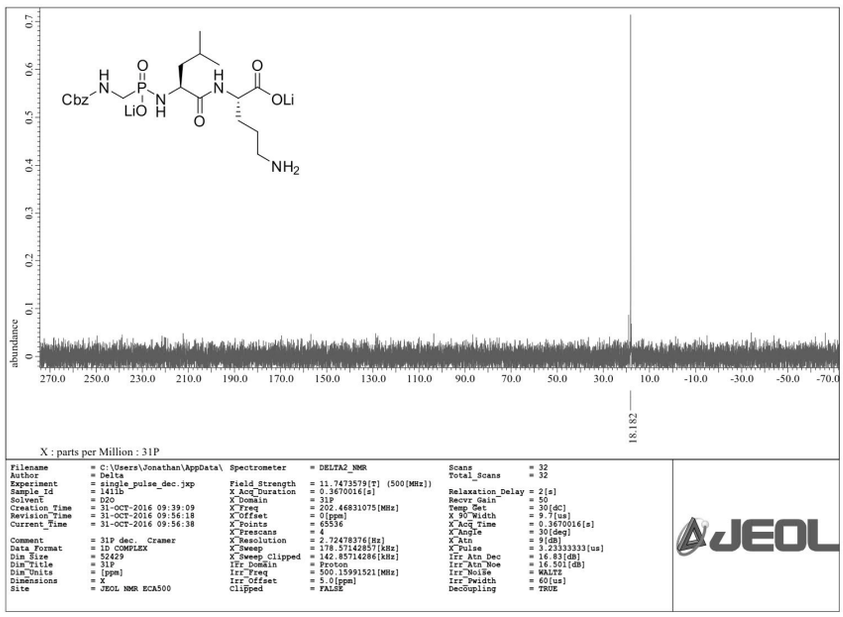


Figure S2.13. ^{31}P NMR Spectrum of **19**.

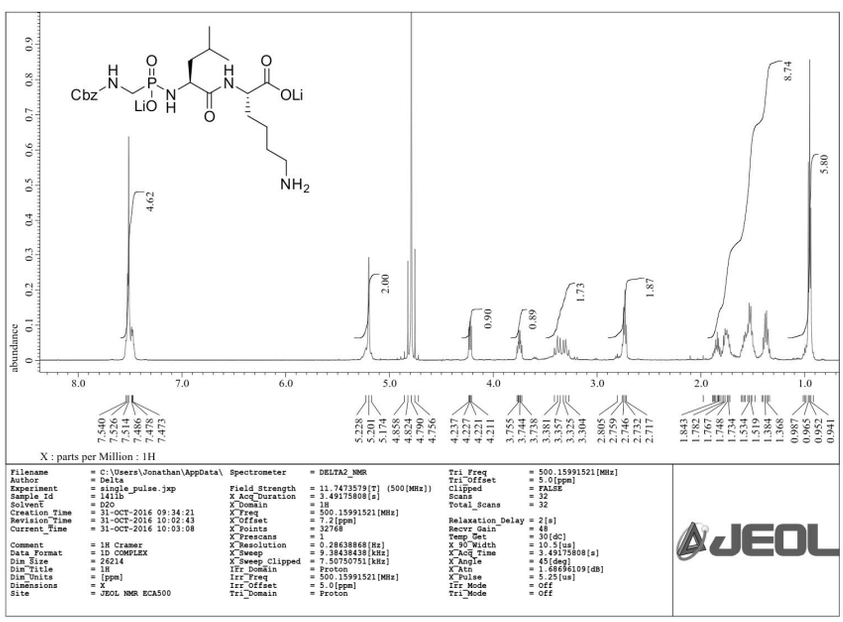


Figure S2.14. ^{1}H NMR Spectrum of **20**.

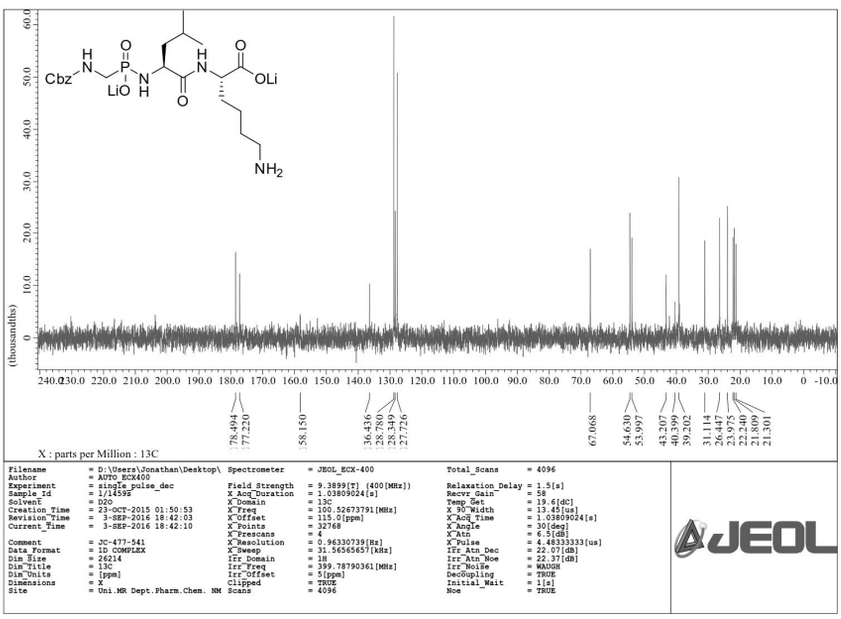


Figure S2.15. ^{13}C NMR Spectrum of **20**.

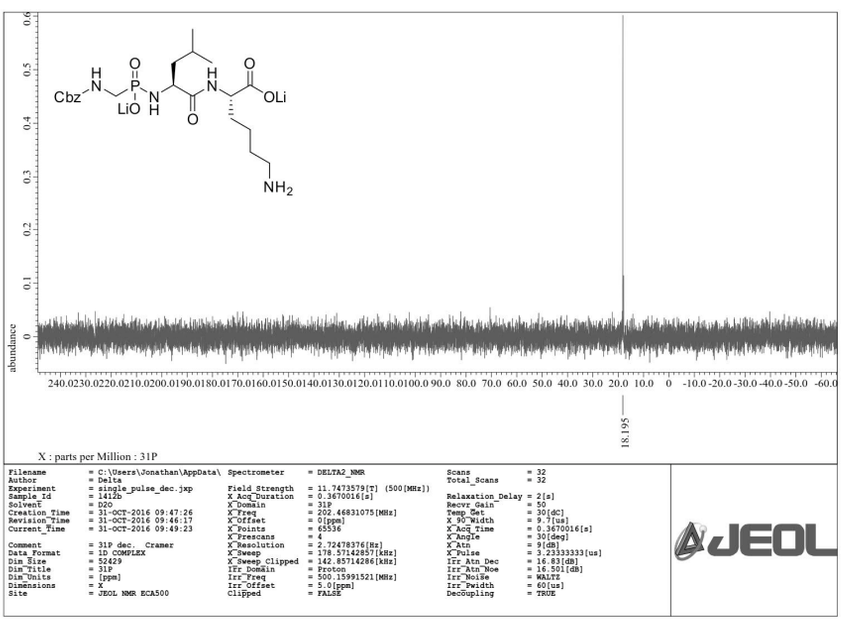


Figure S2.16. ^{31}P NMR Spectrum of **20**.

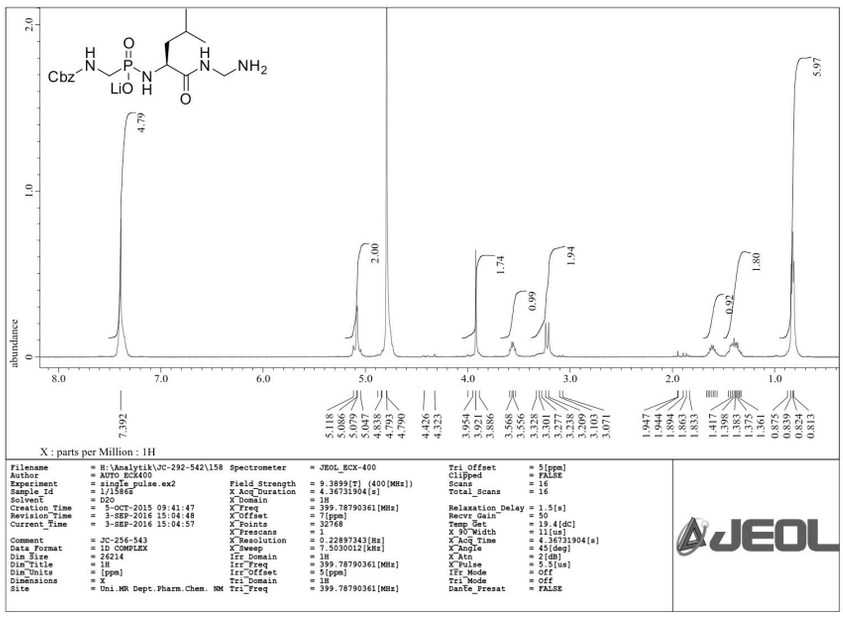


Figure S2.17. ^{1}H NMR Spectrum of **21**.

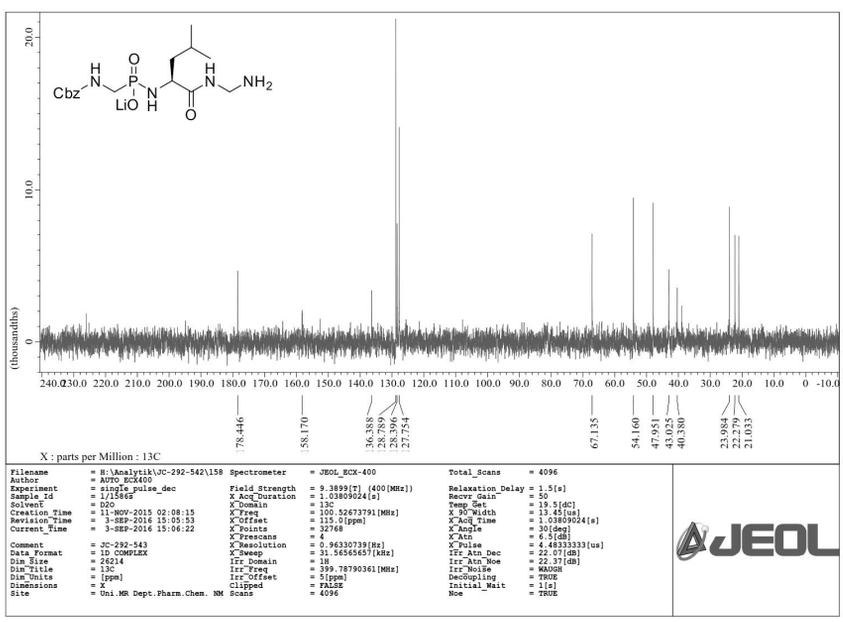


Figure S2.18. ^{13}C NMR Spectrum of **21**.

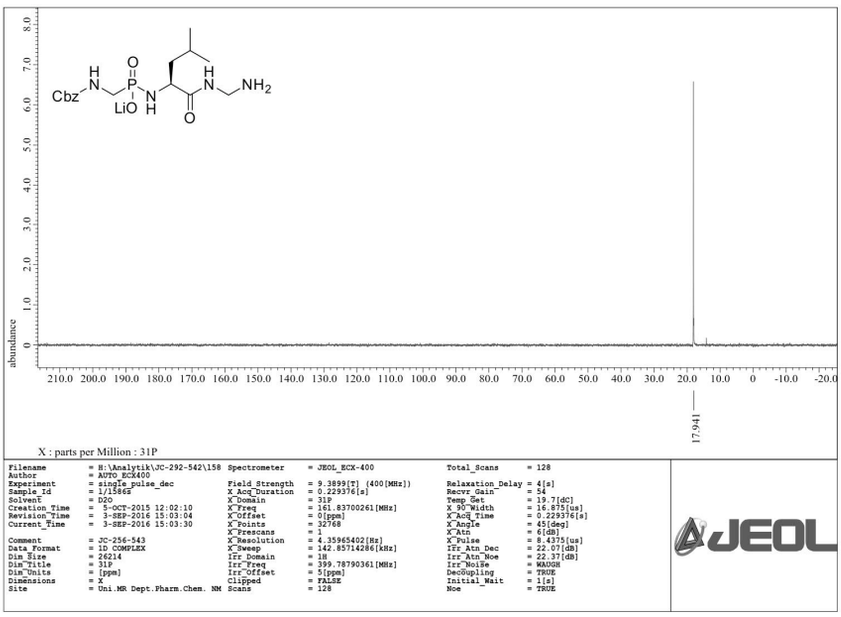


Figure S2.19. ^{31}P NMR Spectrum of **21**.

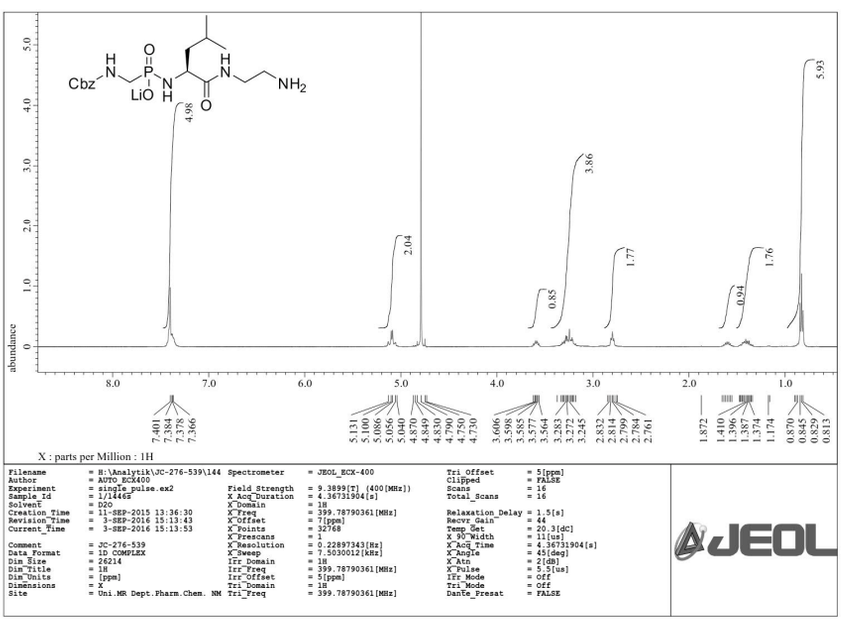


Figure S2.20. ^{1}H NMR Spectrum of **22**.

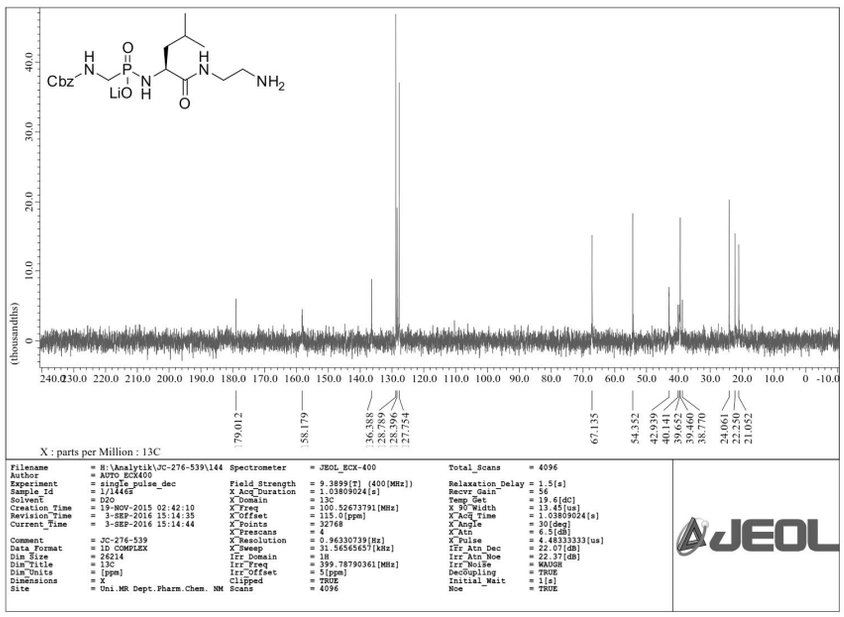


Figure S2.21. ^{13}C NMR Spectrum of **22**.

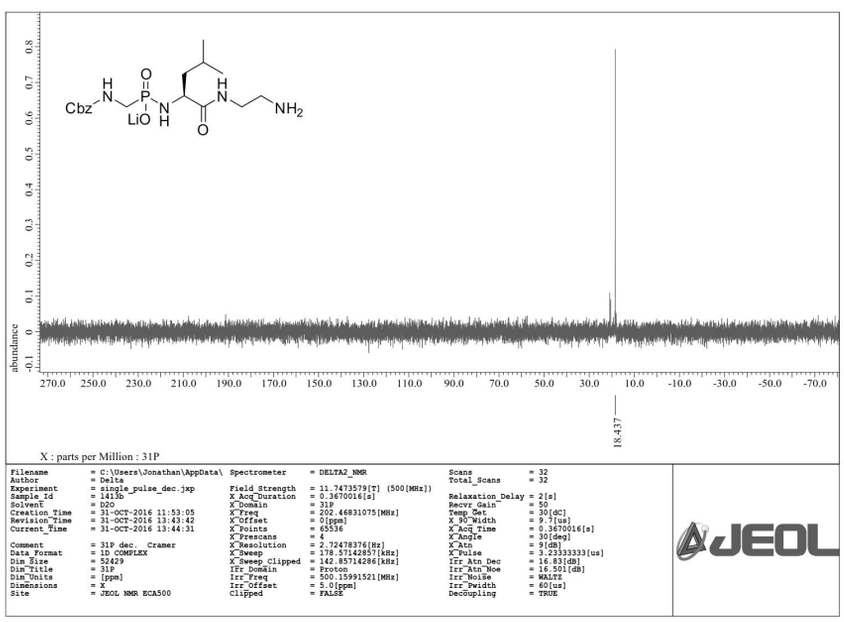


Figure S2.22. ^{31}P NMR Spectrum of **22**.

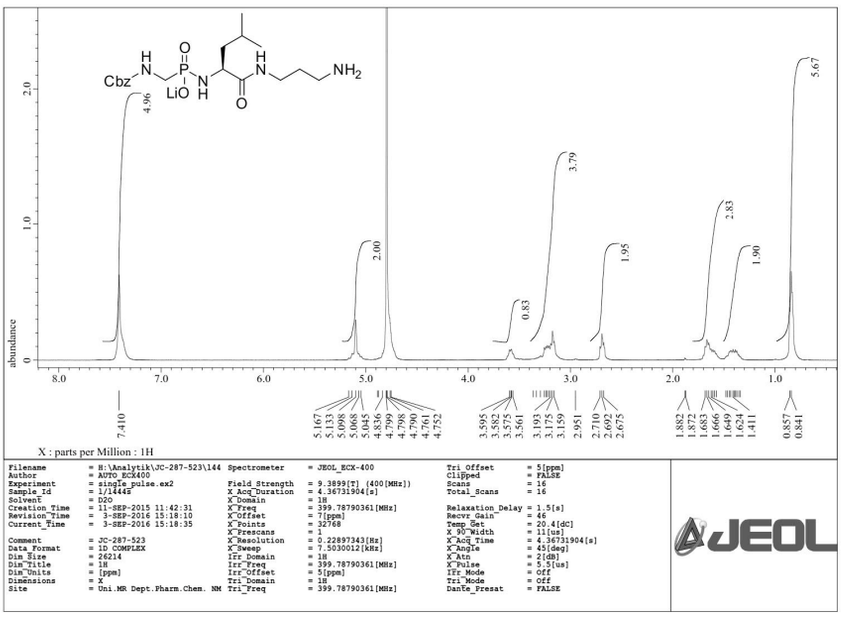


Figure S2.23. ^{1}H NMR Spectrum of **23**.

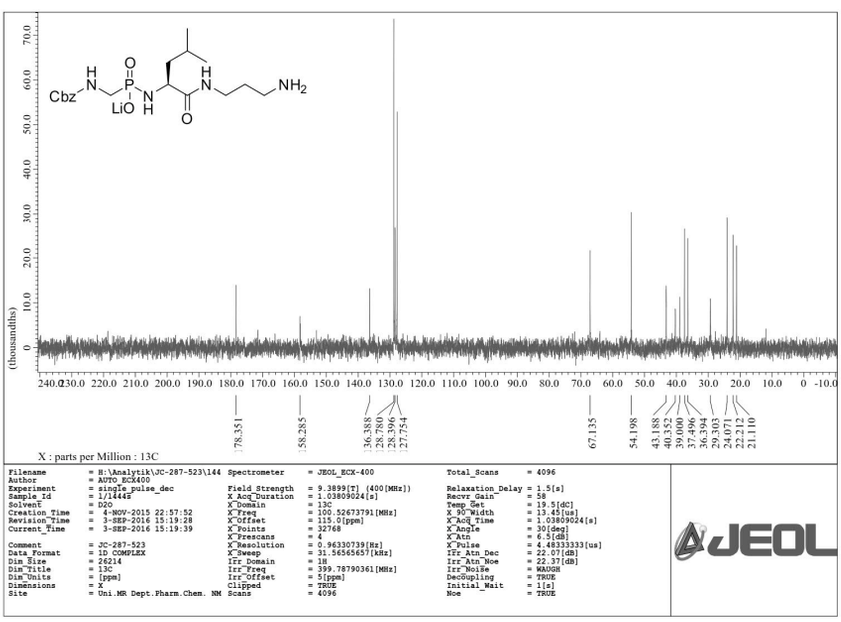


Figure S2.24. ^{13}C NMR Spectrum of **23**.

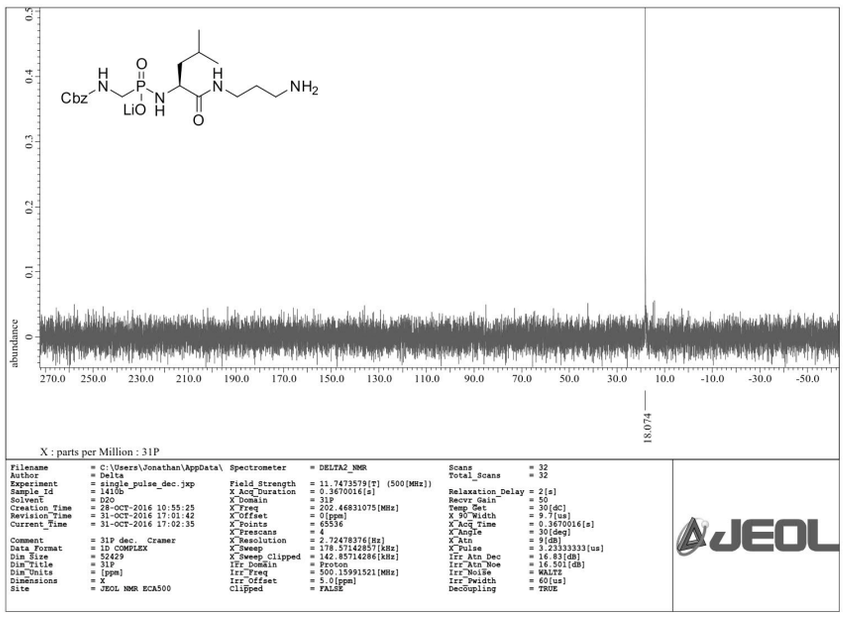


Figure S2.25. ^{31}P NMR Spectrum of **23**.

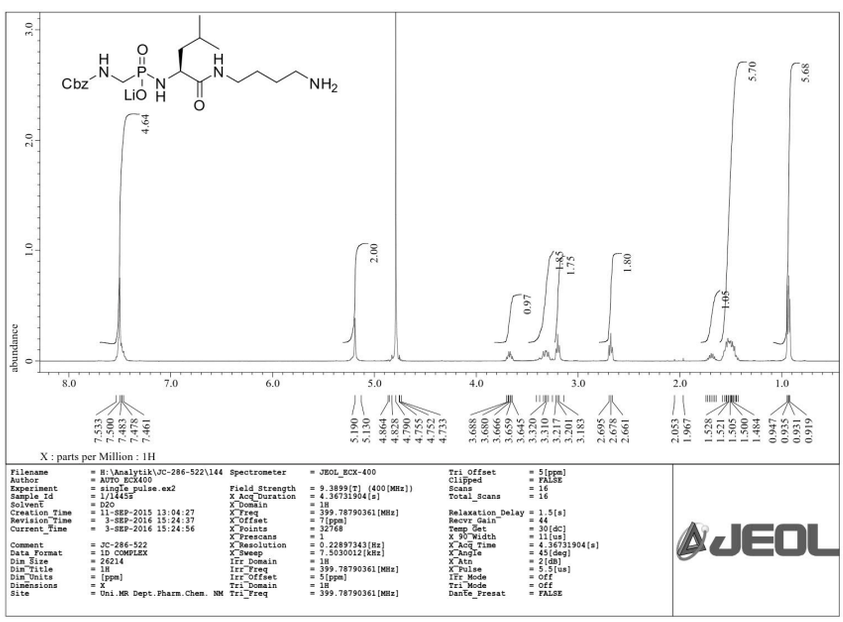


Figure S2.26. ^{1}H NMR Spectrum of **24**.

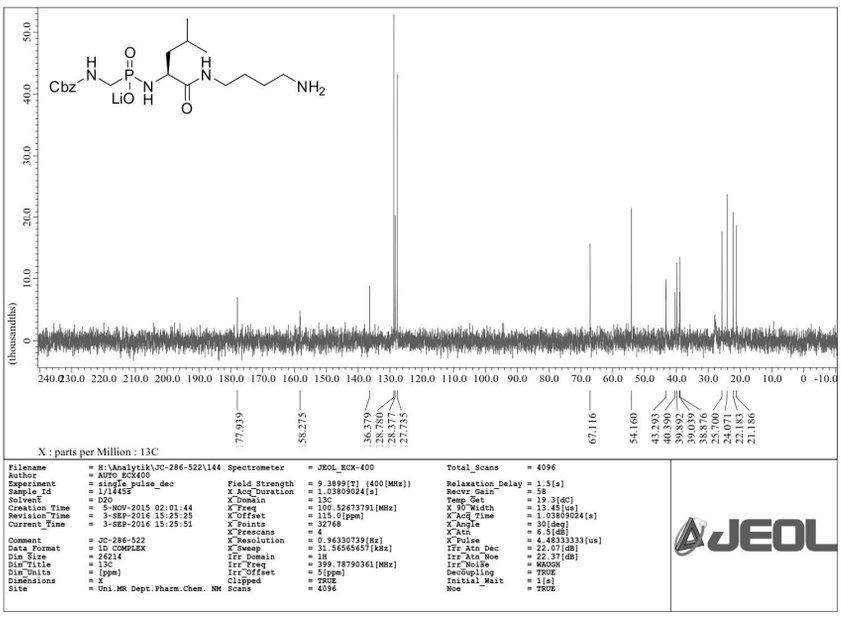


Figure S2.27. ^{13}C NMR Spectrum of **24**.

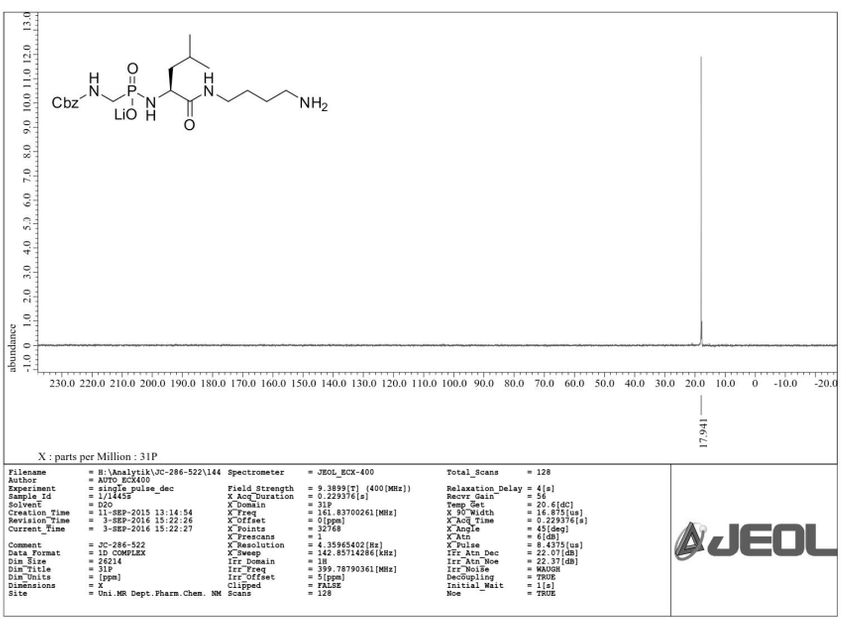


Figure S2.28. ^{31}P NMR Spectrum of **24**.

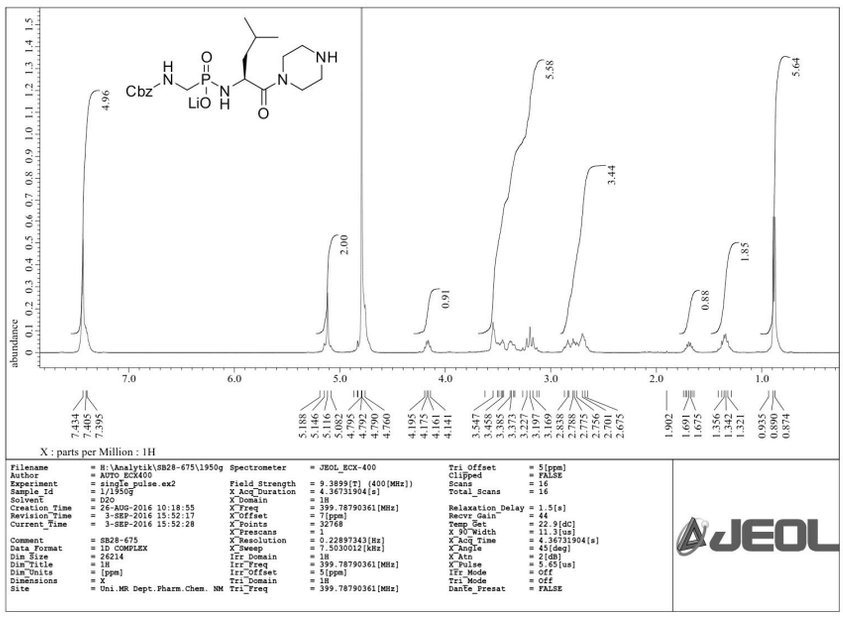


Figure S2.29. ^{1}H NMR Spectrum of **25**.

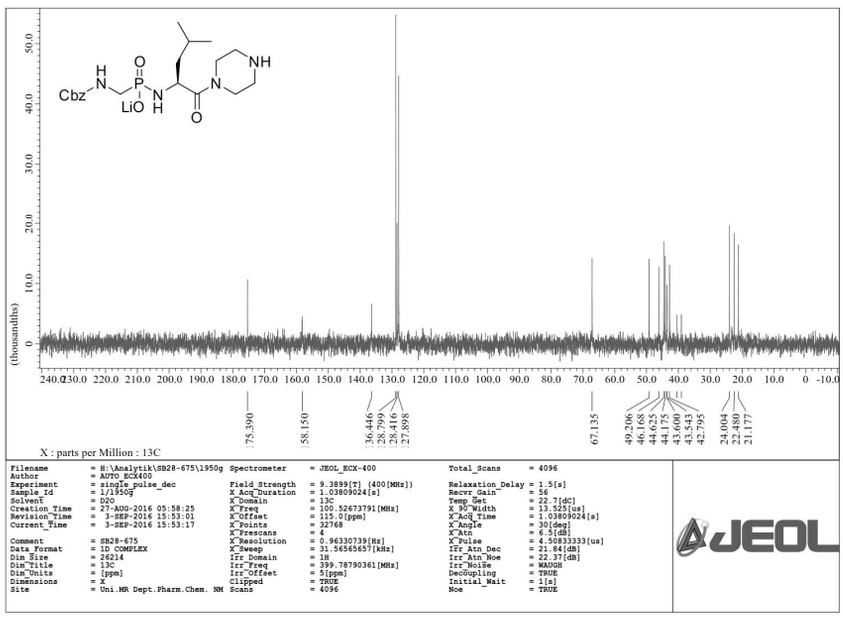


Figure S2.30. ^{13}C NMR Spectrum of **25**.

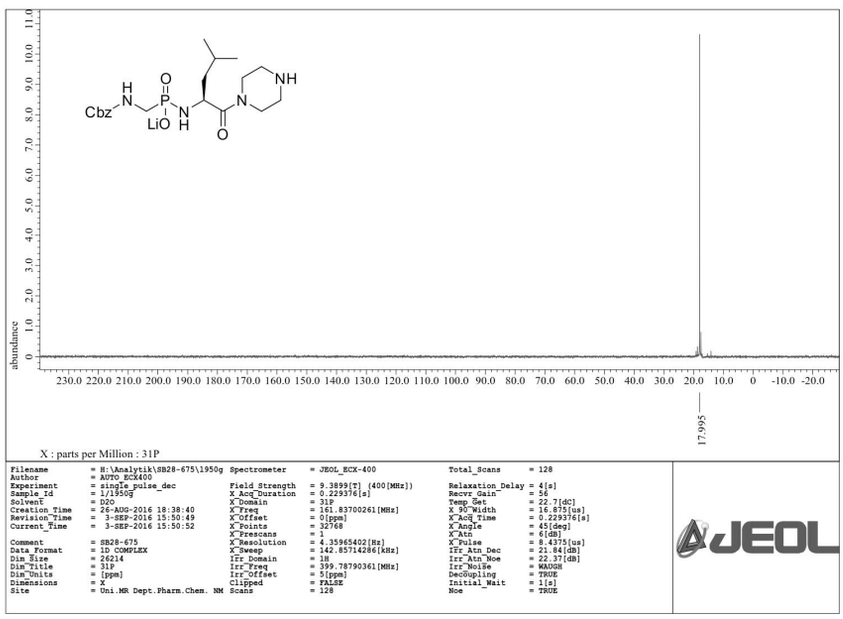


Figure S2.31. ^{31}P NMR Spectrum of **25**.

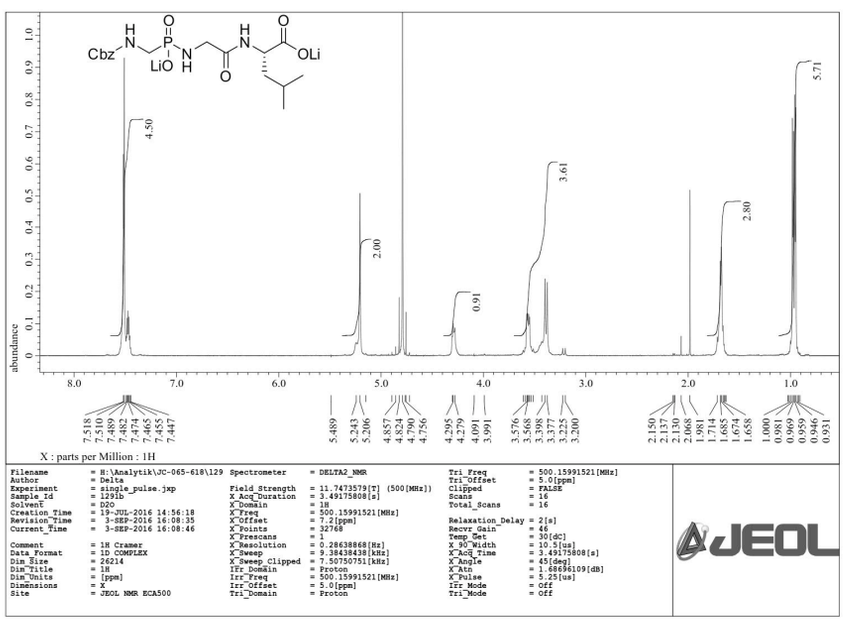


Figure S2.32. ^{1}H NMR Spectrum of **26**.

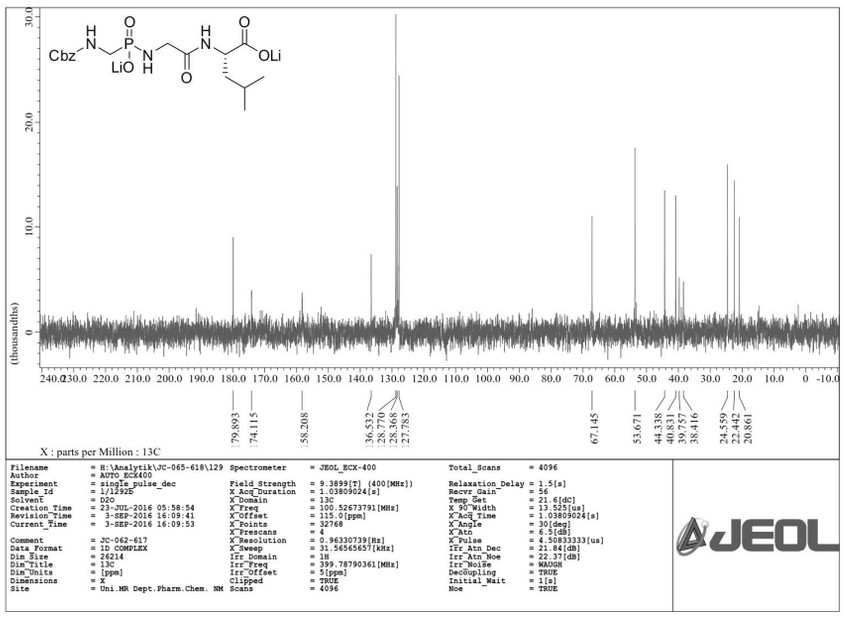


Figure S2.33. ^{13}C NMR Spectrum of **26**.

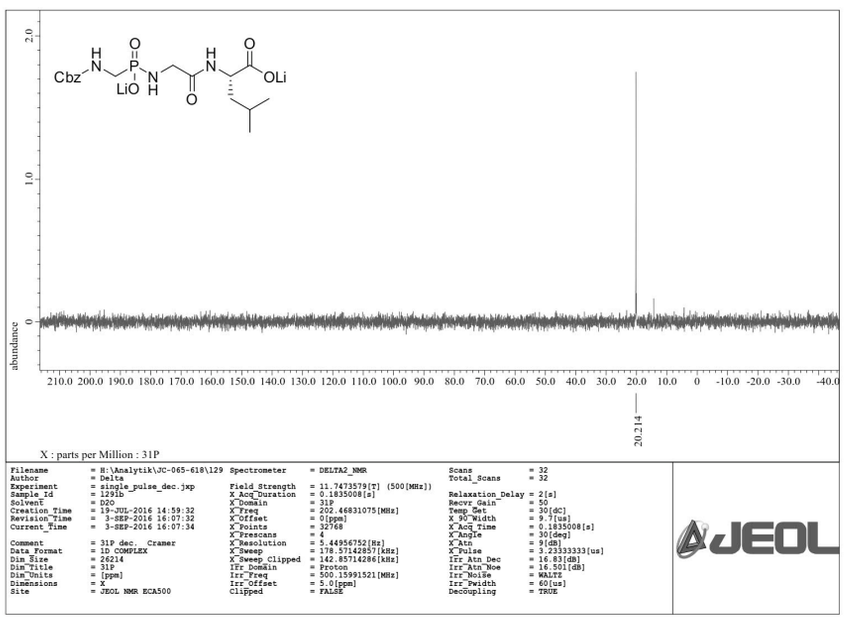


Figure S2.34. ^{31}P NMR Spectrum of **26**.

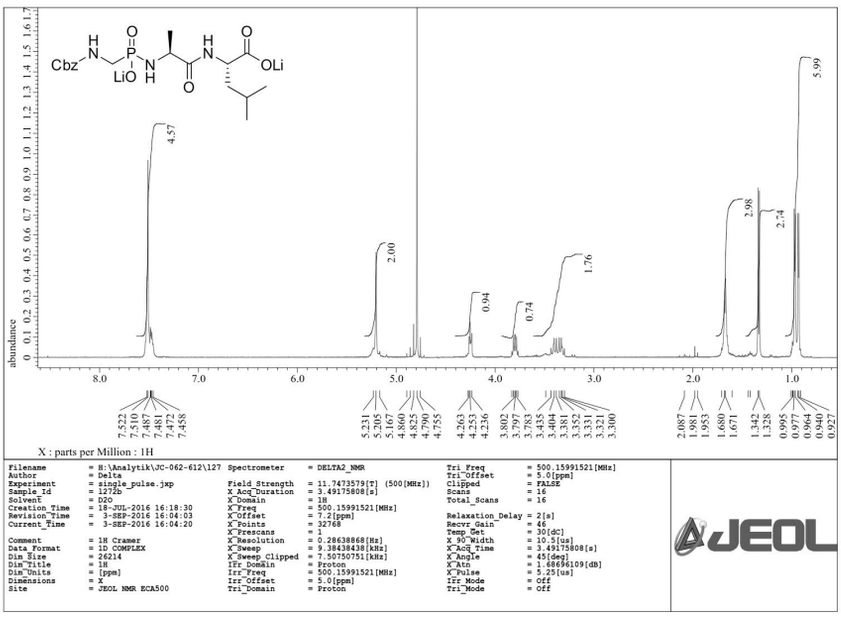


Figure S2.35. ^{1}H NMR Spectrum of **27**.

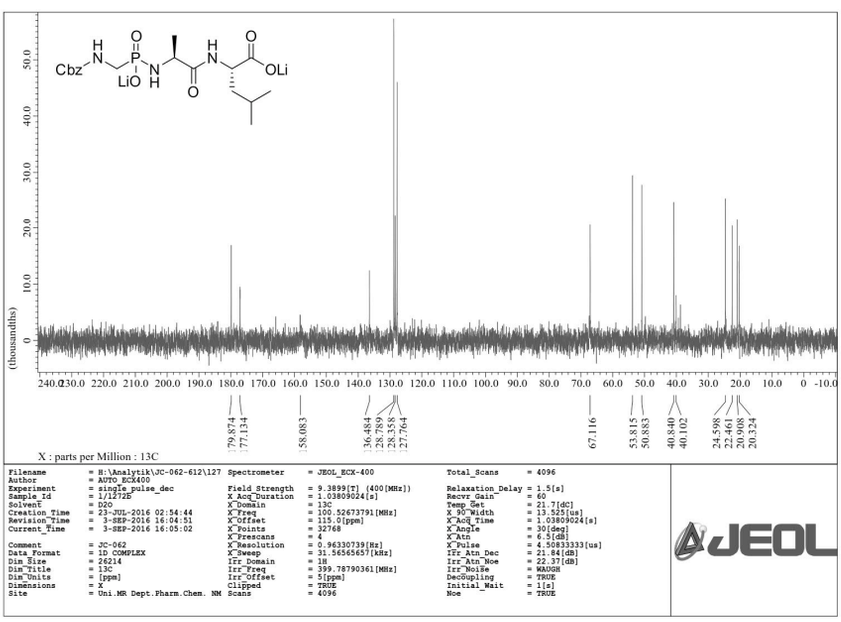


Figure S2.36. ^{13}C NMR Spectrum of **27**.

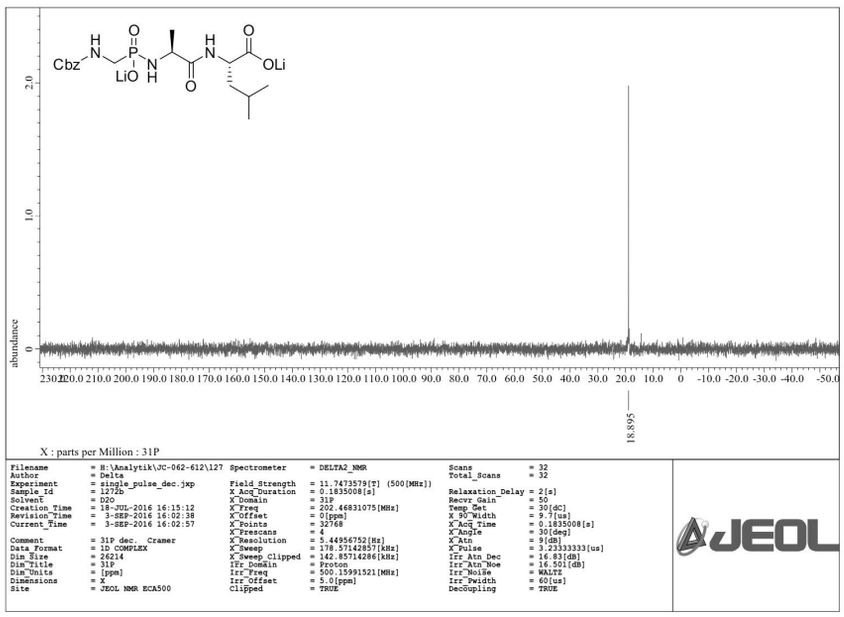


Figure S2.37. ^{31}P NMR Spectrum of **27**.

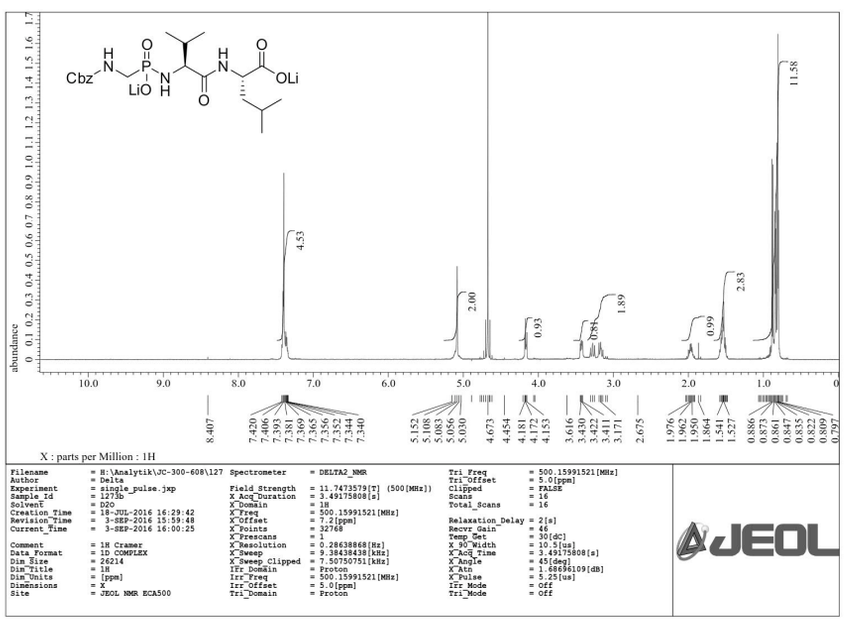


Figure S2.38. ^{1}H NMR Spectrum of **28**.

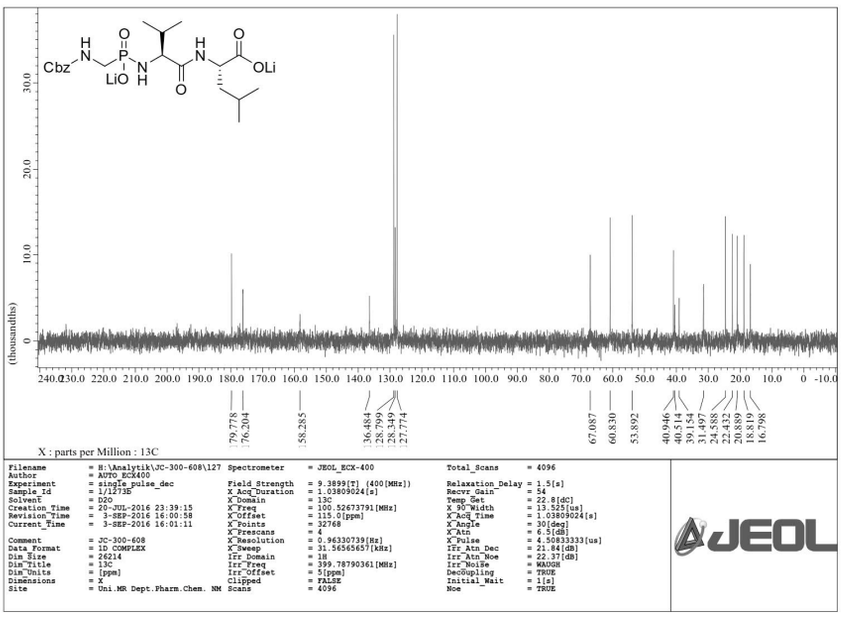


Figure S2.39. ^{13}C NMR Spectrum of **28**.

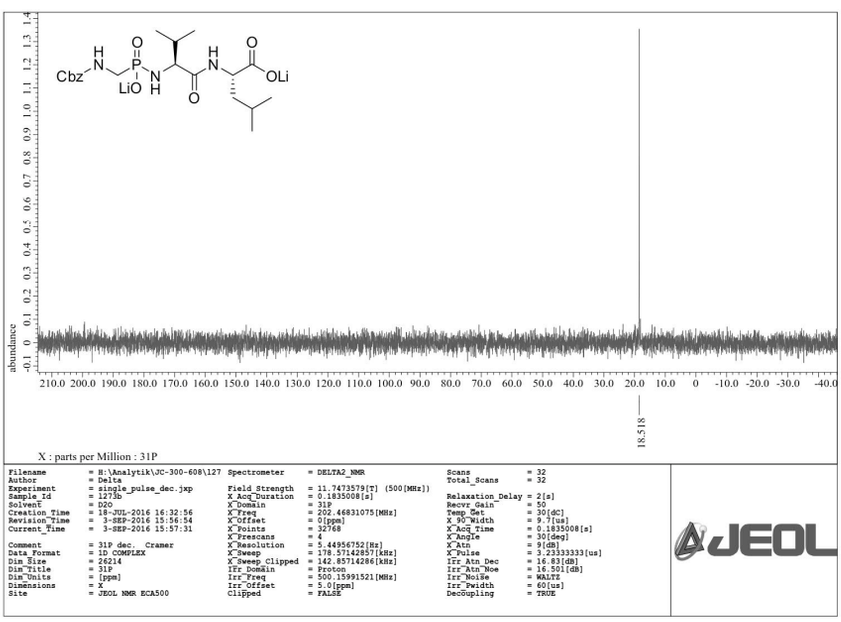
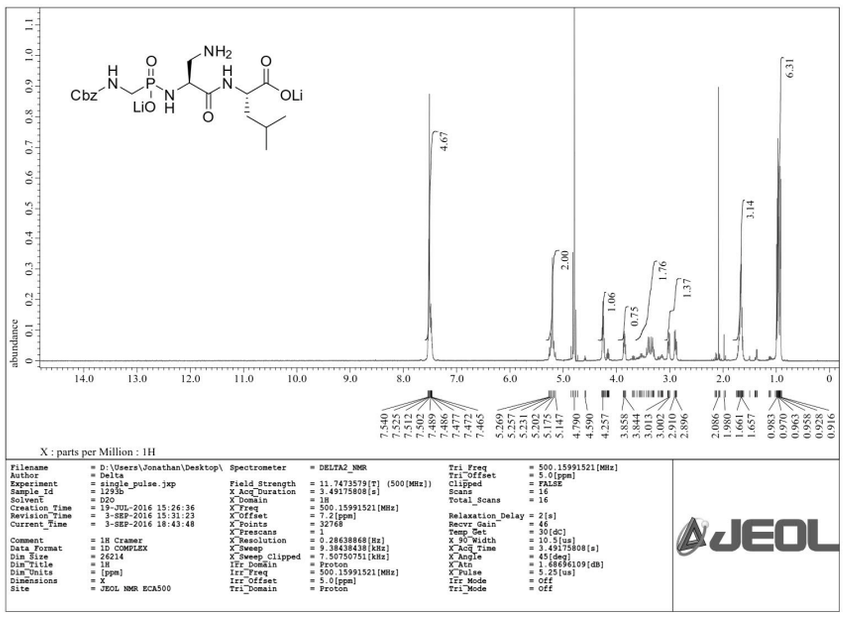
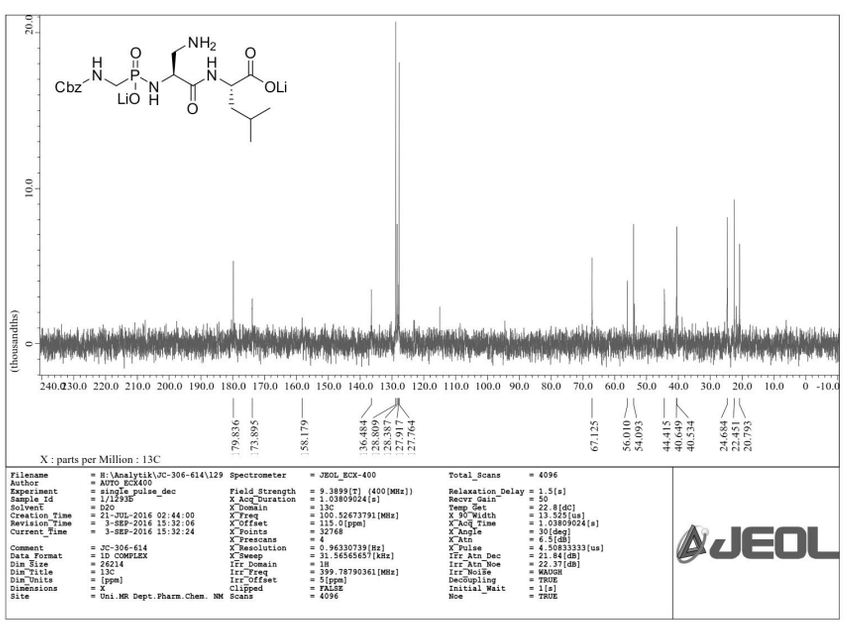


Figure S2.40. ^{31}P NMR Spectrum of **28**.



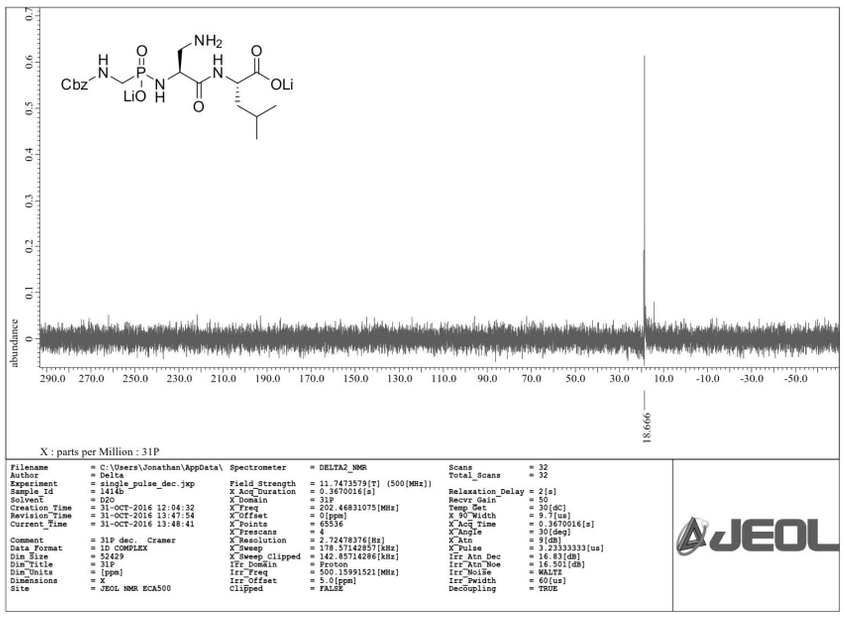
Filename	= D:\Users\Jonathan\Desktop\	Spectrometer	= DELTA2_NMR	Tri_Freq	= 500.15991521[MHz]
Author	= Delta			Tri_Offset	= 5.0[ppm]
Experiment	= single_pulse.jxp	Field_Strength	= 11.7473579[T] (500[MHz])	Clipped	= FALSE
Sample_Id	= 1293b	X_Acq_Duration	= 3.49175808[s]	Scans	= 16
Solvent	= D2O	X_Domain	= 1H	Total_Scans	= 16
Creation_Time	= 19-JUL-2016 15:26:36	X_Freq	= 500.15991521[MHz]		
Revision_Time	= 3-SEP-2016 15:31:23	X_Offset	= 7.2[ppm]	Relaxation_Delay	= 2[s]
Current_Time	= 3-SEP-2016 18:43:48	X_Points	= 32768	Recvr_Gain	= 46
		X_Prescans	= 1	Temp_Set	= 301[dC]
Comment	= 1H Cramer	X_Resolution	= 0.28638868[Hz]	X_90_Width	= 10.5[us]
Data_Format	= 1D COMPLEX	X_Sweep	= 9.38438438[kHz]	X_Acq_Time	= 3.49175808[s]
Dim_Size	= 26214	X_Sweep_Clipped	= 7.50750751[kHz]	X_Angle	= 45[deg]
Dim_Title	= 1H	IFr_Domain	= Proton	X_Atn	= 1.68696109[dB]
Dim_Units	= [ppm]	Irr_Freq	= 500.15991521[MHz]	X_Pulse	= 5.25[us]
Dimensions	= X	Irr_Offset	= 5.0[ppm]	Ifr_Mode	= Off
Site	= JEOL NMR ECA500	Tri_Domain	= Proton	Tri_Mode	= Off

Figure S2.41. ^{1}H NMR Spectrum of **29**.



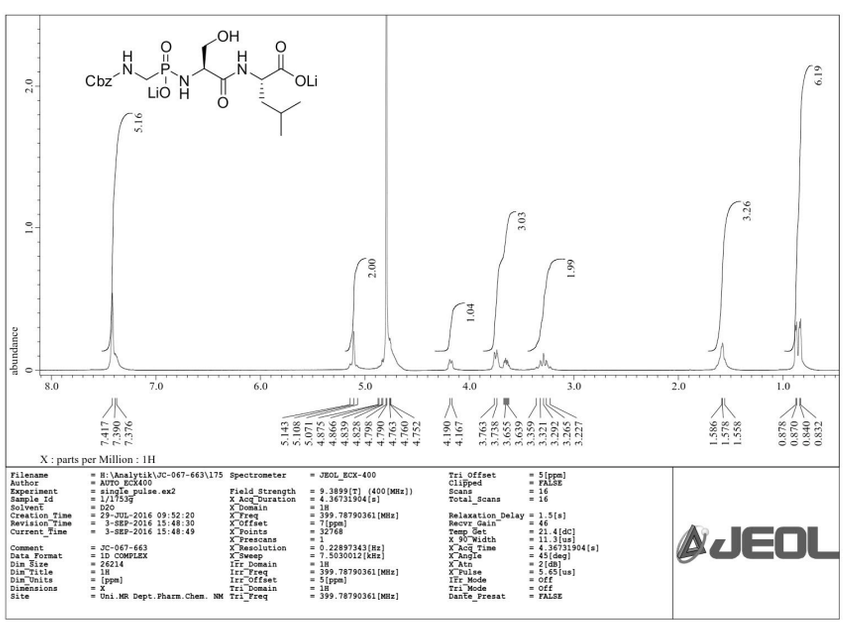
Filename	= H:\Analytik\JC-306-614\129	Spectrometer	= JEOL_ECX-400	Total_Scans	= 4096
Author	= AUTO_ECX400				
Experiment	= single_pulse_dec	Field_Strength	= 9.3899[T] (400[MHz])	Relaxation_Delay	= 1.5[s]
Sample_Id	= 1/1293E	X_Acq_Duration	= 1.03809024[s]	Recvr_Gain	= 50
Solvent	= D2O	X_Domain	= 13C	Temp_Set	= 22.8[dC]
Creation_Time	= 21-JUL-2016 02:44:00	X_Freq	= 100.52673791[MHz]	X_90_Width	= 13.525[us]
Revision_Time	= 3-SEP-2016 15:32:06	X_Offset	= 115.0[ppm]	X_Acq_Time	= 1.03809024[s]
Current_Time	= 3-SEP-2016 15:32:24	X_Points	= 32768	X_Angle	= 30[deg]
		X_Prescans	= 4	X_Atn	= 6.5[dB]
Comment	= JC-306-614	X_Resolution	= 0.96330739[Hz]	X_Pulse	= 4.50833333[us]
Data_Format	= 1D COMPLEX	X_Sweep	= 31.56565657[kHz]	IFr_Atn_Dec	= 21.84[dB]
Dim_Size	= 26214	Ifr_Domain	= 1H	Irr_Atn_Noe	= 22.37[dB]
Dim_Title	= 13C	Irr_Freq	= 399.78790361[MHz]	Irr_NoiSe	= WALTZ
Dim_Units	= [ppm]	Irr_Offset	= 5[ppm]	DecOupling	= TRUE
Dimensions	= X	Clipped	= FALSE	Initial_Wait	= 1[s]
Site	= Uni.MR Dept.Pharm.Chem. NN	Scans	= 4096	Noe	= TRUE

Figure S2.42. ^{13}C NMR Spectrum of **29**.



Filename	= C:\Users\Jonathan\AppData\	Spectrometer	= DELTA2_NMR	Scans	= 32
Author	= Delta			Total_Scans	= 32
Experiment	= single_pulse_dec.jxp	Field_Strength	= 11.7473579[T] (500[MHz])		
Sample_Id	= 1414b	X_Acq_Duration	= 0.3670016[s]	Relaxation_Delay	= 2[s]
Solvent	= D2O	X_Domain	= 31P	Recvr_Gain	= 50
Creation_Time	= 31-OCT-2016 12:04:32	X_Freq	= 202.46831075[MHz]	Temp_Set	= 30[dC]
Revision_Time	= 31-OCT-2016 13:47:54	X_Offset	= 0[ppm]	X_90_Width	= 9.7[us]
Current_Time	= 31-OCT-2016 13:48:41	X_Points	= 65536	X_Acq_Time	= 0.3670016[s]
		X_Prescans	= 4	X_Angle	= 30[deg]
Comment	= 31P dec. Cramer	X_Resolution	= 2.72478376[Hz]	X_Atn	= 9[dB]
Data_Format	= 1D COMPLEX	X_Sweep	= 178.57142857[kHz]	X_Pulse	= 3.23333333[us]
Dim_Size	= 52429	X_Sweep_Clipped	= 142.85714286[kHz]	Ifr_Atn_Dec	= 16.83[dB]
Dim_Title	= 31P	Ifr_Domain	= Proton	Irr_Atn_Noe	= 16.501[dB]
Dim_Units	= [ppm]	Irr_Freq	= 500.15991521[MHz]	Irr_NoiSe	= WALTZ
Dimensions	= X	Irr_Offset	= 5.0[ppm]	Irr_Pwidth	= 60[us]
Site	= JEOL NMR ECA500	Clipped	= FALSE	DecOupling	= TRUE

Figure S2.43. ^{31}P NMR Spectrum of **29**.



Filename	= H:\Analytik\JC-067-663\175	Spectrometer	= JEOL_ECX-400	Tri_Offset	= 5[ppm]
Author	= AUTO_ECX400			Clipped	= FALSE
Experiment	= single_pulse.ex2	Field_Strength	= 9.3899[T] (400[MHz])	Scans	= 16
Sample_Id	= 1/1753g	X_Acq_Duration	= 4.36731904[s]	Total_Scans	= 16
Solvent	= D2O	X_Domain	= 1H		
Creation_Time	= 29-JUL-2016 09:52:20	X_Freq	= 399.78790361[MHz]	Relaxation_Delay	= 1.5[s]
Revision_Time	= 3-SEP-2016 15:48:30	X_Offset	= 7[ppm]	Recvr_Gain	= 46
Current_Time	= 3-SEP-2016 15:48:49	X_Points	= 32768	Temp_Set	= 21.4[dC]
		X_Prescans	= 1	X_90_Width	= 11.3[us]
Comment	= JC-067-663	X_Resolution	= 0.22897343[Hz]	X_Acq_Time	= 4.36731904[s]
Data_Format	= 1D COMPLEX	X_Sweep	= 7.5030012[kHz]	X_Angle	= 45[deg]
Dim_Size	= 26214	Ifr_Domain	= 1H	X_Atn	= 2[dB]
Dim_Title	= 1H	Irr_Freq	= 399.78790361[MHz]	X_Pulse	= 5.65[us]
Dim_Units	= [ppm]	Irr_Offset	= 5[ppm]	Ifr_Mode	= Off
Dimensions	= X	Tri_Domain	= 1H	Tri_Mode	= Off
Site	= Uni.MR Dept.Pharm.Chem. NN	Tri_Freq	= 399.78790361[MHz]	DanTe_Presat	= FALSE

Figure S2.44. ^{1}H NMR Spectrum of **30**.

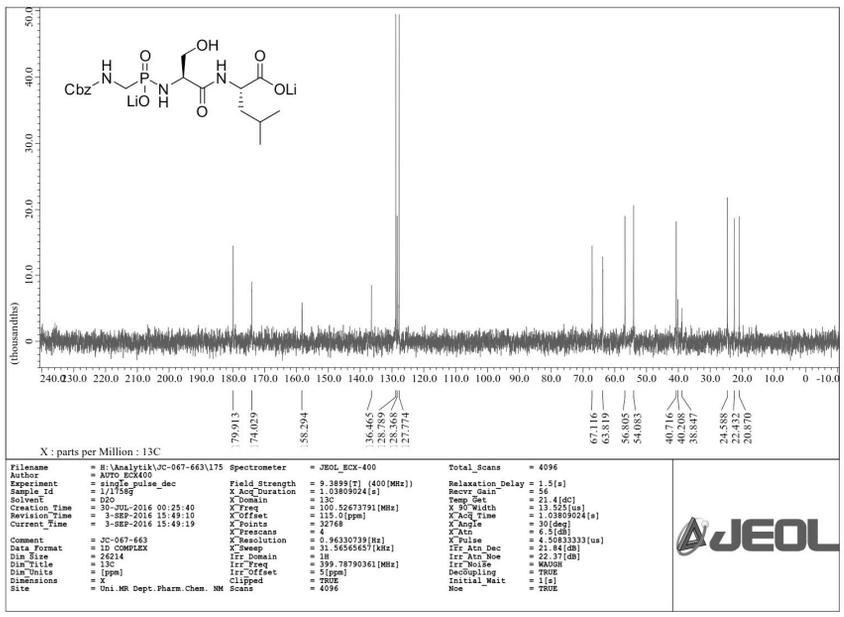


Figure S2.45. ^{13}C NMR Spectrum of **30**.

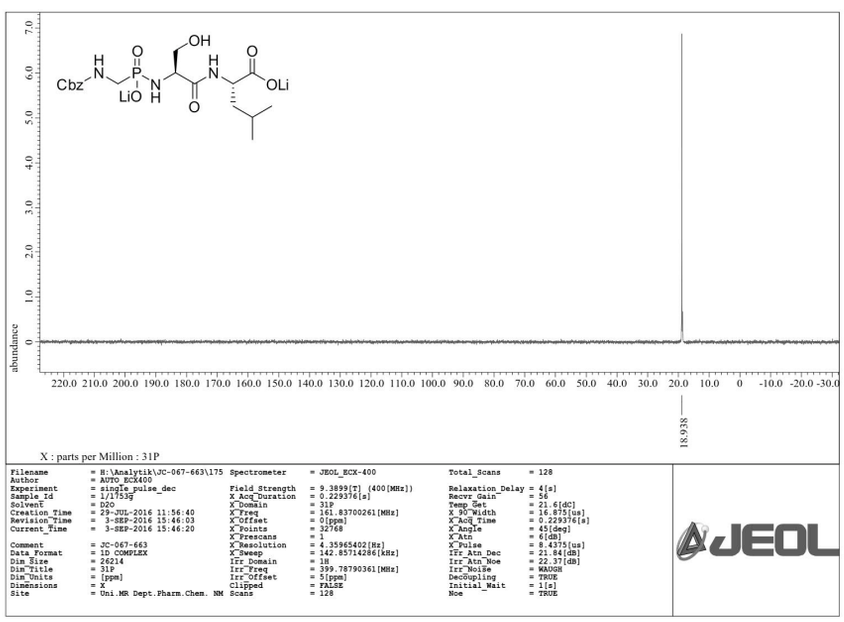


Figure S2.46. ^{31}P NMR Spectrum of **30**.

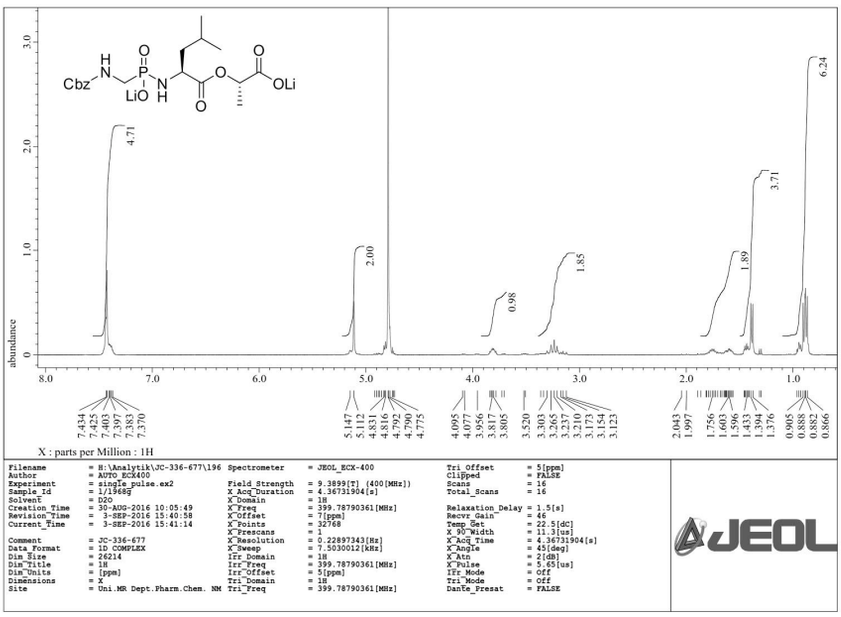


Figure S2.47. ^{1}H NMR Spectrum of **31**.

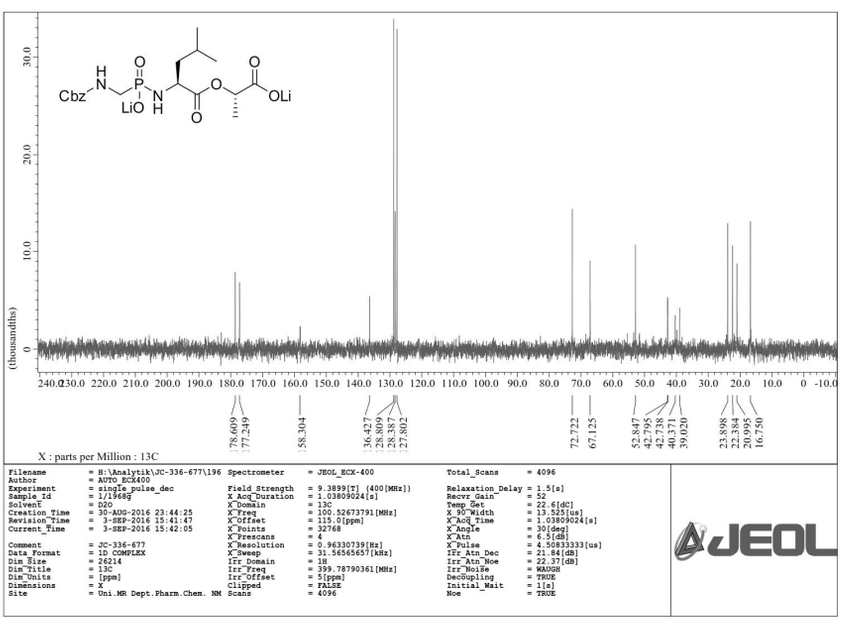


Figure S2.48. ^{13}C NMR Spectrum of **31**.

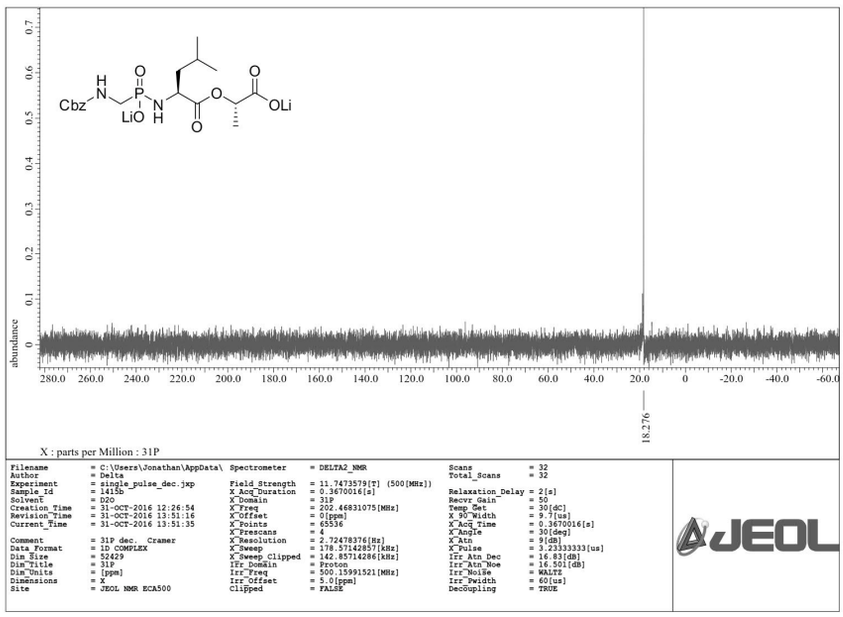


Figure S2.49. ^{31}P NMR Spectrum of **31**.

Chapter 3

Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein-Bound Ligands

Stefan G. Krimmer,[1,*] Jonathan Cramer,[1,*] Michael Betz,[1]
Tobias Wulsdorf,[1] Veronica Fridh,[2] Robert Karlsson,[2]
Andreas Heine,[1] Gerhard Klebe[1]

*S.G.K and J.C. contributed equally to this work

[1]Department of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany

[2]GE Healthcare Bio-Sciences AB, SE-751 84 Uppsala, Sweden

Journal of Medicinal Chemistry **2016**, 59, 10530–10548

3.1 Abstract

A previously studied congeneric series of thermolysin inhibitors addressing the solvent-accessible S_2' pocket with different hydrophobic substituents showed modulations of the surface water layers coating the protein-bound inhibitors. Increasing stabilization of water molecules resulted in enthalpically more favorable binding signature, overall enhancing affinity. Based on this observation, we optimized the series by designing tailored P_2' substituents to improve and further stabilize the surface water network. MD simulations were applied to predict the putative water pattern around the bound ligands. Subsequently, the inhibitors were synthesized and characterized by high-resolution crystallography, microcalorimetry and surface plasmon resonance. One of the designed inhibitors established the most pronounced water network of all inhibitors tested so far, composed of several fused water polygons, and showed 50-fold affinity enhancement with respect to the original methylated parent ligand. Notably, the inhibitor forming the most perfect water network also showed significantly prolonged residence time compared to the other tested inhibitors.

Keywords

protein–ligand interactions, thermodynamic optimization, hydrophobic effect, solvent reorganization, water network prediction, molecular dynamics simulations, X-ray crystallography, isothermal titration calorimetry, surface plasmon resonance, binding kinetics

3.2 Introduction

Structure-based drug design (SBDD) seeks to optimize ligand binding with respect to a given target protein. Thermodynamic parameters such as ΔG, ΔH, $-T\Delta S$, ΔC_p and of binding kinetic properties, such as kinetic association (k_{on}) and dissociation (k_{off}) rate constants, are considered to rationalize and accelerate affinity optimization by a better characterization of the protein–ligand binding process [1,2,24,107–110]. Thermodynamic profiling, at its best based on isothermal titration calorimetry (ITC) [17,111,112], is supposed to provide insights into the molecular interactions determining the affinity of a ligand to its target. However, past experience has shown that optimization and prioritization of compounds guided by thermodynamics is difficult, since the enthalpy–entropy profile reflects the binding event as a whole. Hence, the high complexity of this event largely impedes factorization into individual contributions to binding and often only succeeds for congeneric ligand series with minor structural variations [5,113,114]. The correlation of structural properties with binding kinetic data is presently poorly understood and reasonable correlations have only been established for a limited number of cases [115]. Furthermore, we still face an incomplete comprehension of the fundamental relationships between thermodynamics, kinetics and molecular interactions [115–117], which can lead to false predictions made under overly simplified assumptions. Especially, the impact of the versatile and ubiquitously present water molecules is hardly understood. The involvement of water molecules is a major cause for the inherent complexity, particularly arising from the ability of water to establish H-bonds and the resultant tendency to arrange in differently ordered structures [118]. Water molecules can actively mediate H-bonds across the binding interface between protein and ligand and thereby improve affinity [119]. The displacement of well-ordered water molecules from apolar surfaces is discussed as the driving force of the hydrophobic effect (and *not* the formation of hydrophobic interactions), a process of utmost importance for molecular recognition and drug action. According to the so-called "classical" hydrophobic effect [5,92,120–123], the binding of an apolar ligand portion to a hydrophobic protein cavity correlates with an entropic advantage due to the displacement of well-ordered water molecules into the bulk water phase. In contrast, a "non-classical" hydrophobic effect has been defined, which is enthalpy driven. It has been related to a suboptimal hydration of a protein cavity prior to ligand binding [10,124,125]. In this case, the enthalpy gain upon binding results from the displacement of orientationally mobile and thus entropically favored water molecules into the bulk phase where they can establish better H-bonds than previously observed in the protein cavity.

Despite the popular binary classification of "classical" and "non-classical", the hydrophobic effect can range from entirely entropy-driven to entirely enthalpy-driven. This is determined by the thermodynamic properties of the water molecules that are involved before and after binding and by the ligand and the binding site (especially by their molecular shape and

polarity), which affect the formation of adjacent water networks [14]. Of utmost importance for the thermodynamic signature is the overall inventory of water molecules with respect to their release into or recruitment from the bulk water phase, and the embedding of water molecules into H-bonding networks of varying completeness and perfection in the protein-bound state. In particular, the way water molecules are able to rearrange and establish H-bonding networks around the newly formed solvent-exposed surface of the protein–ligand complex seems to have a significant impact on the thermodynamic binding signature [11,15,93,94,126].

In a previous study [15], we structurally and thermodynamically analyzed a series of congeneric thermolysin (TLN) inhibitors with a peptidomimetic parent scaffold (Figure 3.1A) decorated with different hydrophobic P_2' substituents (Figure 3.1B). We selected the zinc metalloprotease TLN from *Bacillus thermoproteolyticus* for our studies [34,127,128], as this enzyme has been frequently used as class representative and exhibits excellent crystallographic properties [36,129]. In addition, TLN is quite rigid because of its high thermal stability, thus reducing structural adaptions and facilitating comparative analyses. The active site of TLN is composed of three subpockets (Figure 3.1A). Firstly, the rather unspecific S_1 pocket, a hydrophobic cavity that recognizes aromatic portions such as Phe. Secondly, the S_1' specificity pocket, a predominantly hydrophobic, deep and narrow pocket, which preferentially accommodates hydrophobic amino acids such as Val, Leu, Ile and Phe, and thirdly the S_2' subsite, a hydrophobic, flat, bowl-shaped pocket, which is well-accessible to bulk water molecules [127]. We selected the S_2' pocket for our studies. Within a previously investigated series [15], we increased the size of the P_2' substituents addressing the well-solvated S_2' pocket systematically from a sole methyl to a phenylethyl substituent as displayed in Figure 3.1B. By detailed ITC analyses (Figure 3.1B), we revealed a difference in binding affinity (K_d: dissociation constant of the thermodynamic equilibrium) of more than one order of magnitude [15], or expressed as standard Gibbs free energy, a maximum $\Delta\Delta G°$ of 7.0±0.4 kJ mol^{-1}. Remarkably, $\Delta\Delta G°$ factored in a huge enthalpy–entropy variation, and indicated pronounced enthalpy–entropy compensation, a phenomenon frequently observed in drug optimization [4,121,130,131]. We could correlate the observed variations in the thermodynamic profiles with crystallographically observed structural changes of the P_2' substituents and, triggered thereby, modulations of the adjacent surface water solvation layer [15]. The $\Delta H°$ contributions appeared to be more favorable (and $-T\Delta S°$ less favorable) where a better-ordered water network was established next to the surface of the newly formed protein–ligand complex. In contrast, $-T\Delta S°$ apparently increased (and $\Delta H°$ decreased) where the first solvation layer next to the bound ligand was unfavorably disrupted. Furthermore, the increasing contribution to desolvate the gradually growing P_2' substituents seemed to

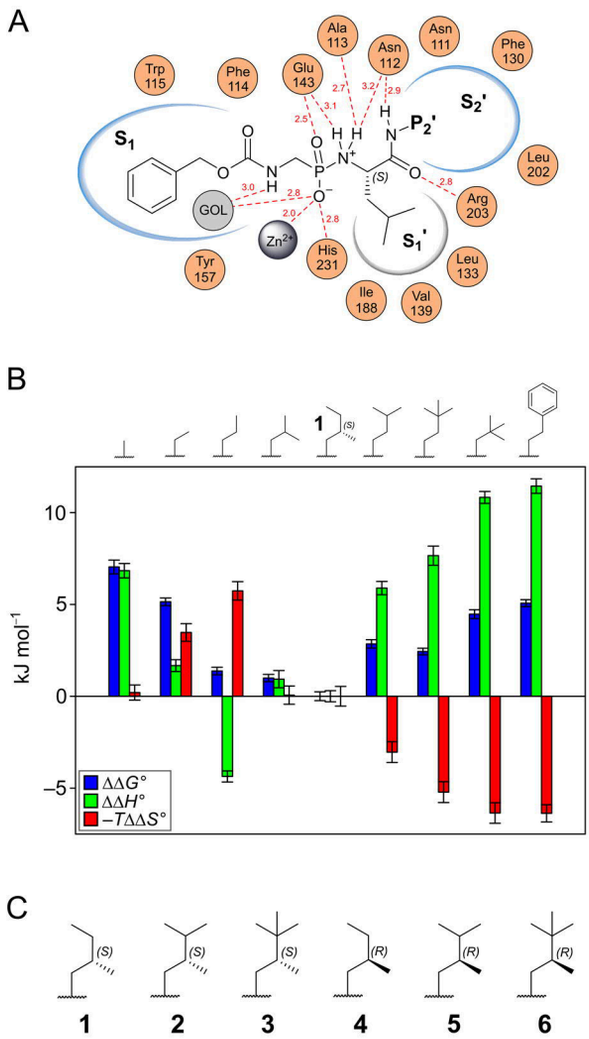


Figure 3.1. (A) Schematic binding mode and protonation state of the parent ligand scaffold in complex with TLN addressing the active site and the S_1, S_1' and S_2' subpockets. Structural components of TLN determinant for the shape of the binding pocket are indicated (GOL = glycerol). Solvent accessibility of sub-pockets is indicated by a blue contour region. H-bonds of at least moderate strength (≤3.2 Å) between parent scaffold and TLN are indicated by red dotted lines, all established to the side-chains of the respective amino acids with the exception of Ala113 (H-bond to backbone carbonyl O). The carboxyl function of Glu143 is protonated while in complex with the ligand and thus can form an H-bond to the double-bonded phosphonamidate O of the ligand [93]. **(B)** Differences in the thermodynamic profiles of the ligands from the previous study [15] were compared to **1** from the latter study. This ligand was also used as a starting point for the design of the ligand series of the current study (**2**–**6**). A positive term resembles a less favorable parameter of $\Delta H°$, $-T\Delta S°$ or $\Delta G°$ relative to ligand **1**, whereas a negative term resembles a more favorable parameter. The chemical structures of the P_2' groups of the ligands are displayed, and the parent scaffold of them is displayed in Figure 3.1A. The graph was prepared with the thermodynamic parameters as analyzed earlier [15]. Mean standard deviations were calculated by the square root of the sum of variances. **(C)** P_2' substituents of phosphonamidate-type TLN inhibitors **1**–**6**. The parent scaffold of the ligands is identical as displayed in Figure 3.1A.

enhance binding entropy, as expected for the "classical" hydrophobic effect of a well-hydrated, apolar cavity [132]. Strikingly, the inhibitor with the highest affinity (ligand **1**, Figure 3.1B) showed both, a pronounced burial of its relatively large P_2' group along with a well-established surface water network wrapping around this substituent in its protein-bound state [15]. This resulted in a well-balanced thermodynamic profile driven by favorable enthalpic and entropic contributions, overall resulting in an increase in binding affinity. Consequently, optimization of the surface water network wrapping around the partly solvent-exposed P_2' substituent appears an useful approach to enhance ligand binding, since the enthalpic gains seem to overall improve affinity.

In the present study, we want to validate this working hypothesis by systematically improving of the surface water network around a newly formed protein–ligand complex to modulate its thermodynamic binding profile and thus increase affinity of a bound ligand. Starting with the best and already fairly well-optimized ligand **1** of our previous series (Figure 3.1B), we designed five additional ligands (**2**–**6**, Figure 3.1C) based on the carboxybenzyl-Gly-(PO_2)-L-Leu-NH_2-P_2' parent scaffold (Figure 3.1A), and attached distinct apolar P_2' substituents to generate differently shaped solvent-exposed surfaces in complex with TLN. Prior to ligand synthesis, we used molecular dynamics (MD) simulations to predict the quality and completeness of the surface water network established around the newly formed complex [133]. The MD simulations suggested the highest completeness and quality for the complex with **3** and lowest for its epimer **6**. As the designed ligands seem to be promising candidates to validate our hypothesis, we synthesized all five to study the established water networks around the formed complexes by X-ray crystallography and thermodynamically by ITC. Furthermore, as we also expected an impact on the binding kinetic properties, we studied the association and dissociation rate constants k_{on} and k_{off} by surface plasmon resonance (SPR).

3.3 Results

3.3.1 Ligand design and solvation pattern prediction using MD simulations

Our design hypothesis to maximize the desolvation of increasingly bulky P_2' substituents along with an energetically favorable ("optimal") surface water network to enhance binding affinity started with the binding pose of **1**, which was already characterized in a previous study (PDB code 4MZN) [15]. This ligand showed a well-established surface water network toward the left rim of the binding pocket including an energetically favorable five-membered water polygon, interconnected by H-bonds (Figure 3.2A, left panel). Deficiencies in the network are suggested on the lower and right-hand side of the S_2' pocket (direction relative to the view of the figure). Here, the *tert*-butyl portion of the ligand with the 2,2-dimethylbutanyl P_2' group (Figure 3.2A, right panel), exhibiting two additional terminal methyl groups relative to **1**, stabilized a more

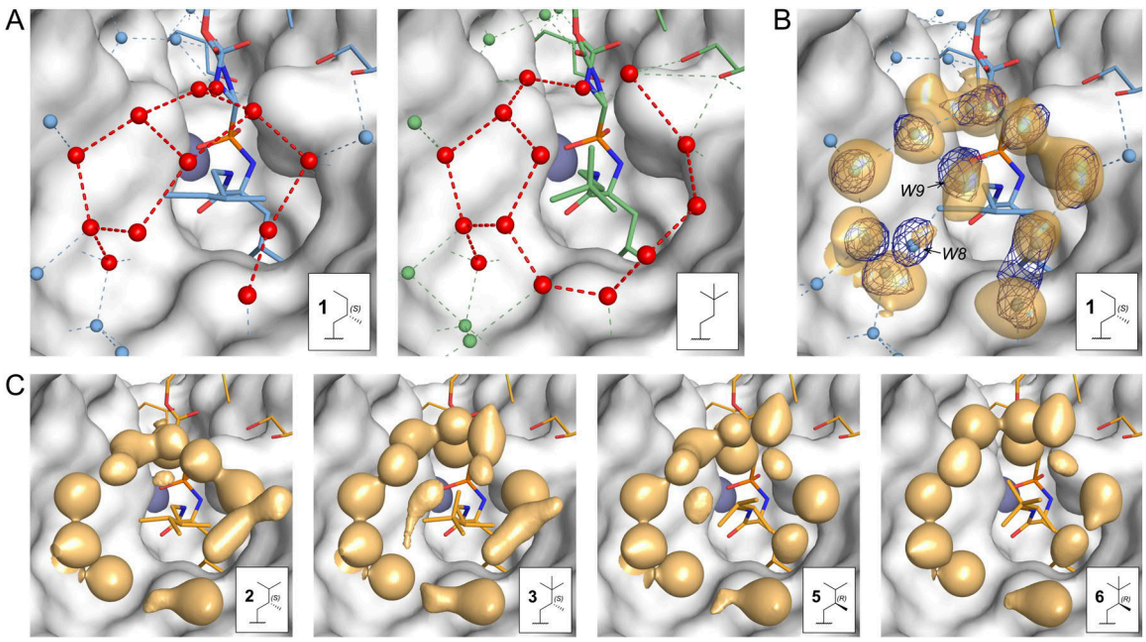


Figure 3.2. Ligand design for optimized surface water network stabilization and evaluation of solvation sites by MD simulations. The inlets of the figures represent the chemical structures of the P_2' substituents of the displayed ligands, their parent scaffold is shown in Figure 3.1A. In all panels, the solvent excluded surface of TLN is displayed in white. **(A)** Crystal structure of **1** (blue) and of the TLN ligand with the 2,2-dimethylbutanyl P_2' group (green). Water molecules are displayed as spheres, H-bond distances between them as dotted lines. Water molecules from the first solvation layer of the P_2' groups are highlighted in red. **(B)** Comparison between the crystal structure of TLN-**1** and the solvation-site predictions by MD simulations. The crystallographically determined binding modes of ligand and additive molecules also used in the MD simulations are displayed as blue stick models with color-coded heteroatoms. Water molecule positions determined in the crystal structure are displayed as blue spheres, and the F_o–F_c omit electron density is displayed as dark blue mesh (contour level 3σ) for the water molecules in the first solvation layer of the P_2' group. H-bond distances are indicated as blue dotted lines. Positions of water molecules, which are discussed in the main text, are labeled with identifiers according to Figure 3.4. The yellow, semitransparently contoured regions show computed areas in the first solvation layer of the P_2' groups with an occupancy probability by a water molecule of at least 48%. **(C)** Hydration sites of the designed ligands as predicted by MD simulations. The modeled coordinates for ligand and additive molecules used in the MD simulations are displayed as yellow stick models (heteroatoms color-coded). The yellow contoured regions represent areas in the first solvation layer of the P_2' groups with an occupancy probability by a water molecule of at least 48%.

complete network in this region, and also established the favorable pentagonal polygon to the left. However, its water network is unfavorably disrupted on top of the *tert*-butyl portion, resulting in an entropically highly favored system with an overall lowered affinity (Figure 3.1B). We therefore envisioned merging the features of both P_2' substituents into size-increased **2** (comprising only one of the two additional methyl groups of the *tert*-butyl portion) and **3** (exhibiting both additional methyl groups of the *tert*-butyl portion). As our design involved

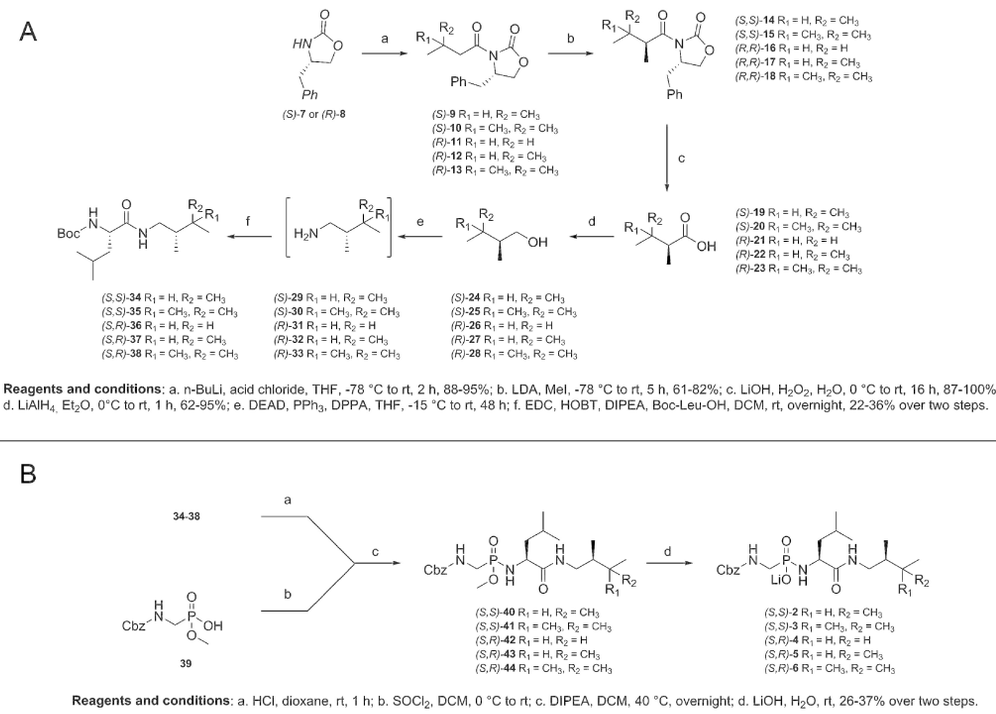
the creation of stereogenic centers, we also considered the epimers **5** and **6** of **2** and **3** in our subsequent MD evaluation.

To predict and analyze the pattern of water solvation sites around the designed P_2' substituents, we applied our recently introduced MD approach [133] to simulate TLN-**1**, TLN-**2**, TLN-**3**, TLN-**5** and TLN-**6**. Ligand **4** (the epimer of **1**, Figure 3.1C) was not considered for the MD simulations, as this ligand was only synthesized at a late stage of the study with the purpose to complete the congeneric series and to further validate the influence of the rearrangement of water molecules on the thermodynamics of protein–ligand binding. This strategy provided the opportunity to validate our MD protocol on TLN-**1**, as we had already determined a high-resolution crystal structure (1.17 Å) for this complex [15]. MD simulations were run for 20 ns and the spatial positions of water molecules were recorded every 2 ps along the trajectory, from which the propensity of water molecules to occupy the indicated solvation sites was calculated. For the novel complexes, the crystal structure of TLN-**1** was used as a template. During the MD simulations, non-hydrogen atoms and non-water molecules were constrained to their starting coordinates. Similar protocols were applied to model the designed complexes with **2**, **3**, **5** and **6** (for further details, see the Experimental Section). For TLN-**1**, the computed results are superimposed with the difference electron densities of the water molecules found in the crystal structure (Figure 3.2B). The displayed solvation sites encompass a probability greater than 48% to record a water molecule along the trajectory at this site. This contour level was adjusted by visual inspection of the computed map to qualitatively match with the contouring of the crystallographically determined F_o–F_c difference electron density at the commonly applied 3σ level. The results matched convincingly well. Only the site *W8* is predicted as less populated compared to the crystal structure and *W9* was suggested as being slightly displaced by the MD approach. Mutually facing the distributions of the computed solvation sites indicates that TLN-**3** displays the most densely packed and complete surface water network in the series (Figure 3.2C). The simulation of TLN-**5** also assigned a rather elaborate surface water network to this complex, whereas TLN-**2** and in particular TLN-**6** show a gap in the water network capping their P_2' substituents.

3.3.2 Stereoselective synthesis of the congeneric phosphonamidate inhibitors 2–6

Stimulated by the simulation results, we decided to synthesize **2**–**6**. Compound **4**, the epimer of **1**, was included for reasons of comparison.

Scheme 3.1 illustrates representatively the synthesis route to prepare **2**–**6**. The stereogenic center in the P_2' portion of the ligands was synthetically accessible by a strategy employing a chiral oxazolidinone auxiliary. 4-Benzyl oxazolidinone (**7** or **8**) was treated with *n*-BuLi



Scheme 3.1. (A) Synthesis of the peptidic P2′ portions of 2–6. (B) Phosphonamidate coupling and deprotection of **34**–**38**.

followed by the respective acid chlorides. In a diastereoselective enolate alkylation of the resulting *N*-acyloxazolidinones (**9**–**13**), intermediates **14**–**18** were synthesized in a diastereomeric ratio of 13:1–16:1. With the exception of **16**, the concentration of the desired diastereomers could be improved to a ratio of >20:1 by recrystallization from cyclohexane. The auxiliary was removed by hydrolysis with $LiOH/H_2O_2$. The chiral carboxylic acids **19**–**23** were subsequently reduced to the corresponding alcohols (**24**–**28**) with $LiAlH_4$. The conversion to the peptidic intermediates **34**–**38** was carried out by a multistep procedure involving a Mitsunobu reaction with DPPA followed by an *in situ* Staudinger reduction of the resulting azides. Due to their highly volatile nature, the intermediate amines **29**–**33** were not isolated. After an aqueous workup procedure, they were reacted with Boc-Leu-OH under standard EDC coupling conditions.

Phosphonic acid monoester **39** was synthesized following a modified literature procedure [134]. For the phosphonamidate coupling reaction, **39** was activated with $SOCl_2$. The peptidic intermediates **34**–**38** were deprotected using a 4 M solution of HCl in dioxane and reacted with

the activated phosphonic acid. The final deprotection of the coupling products **40**–**44** with aqueous LiOH solution followed by semi-preparative HPLC purification afforded inhibitors **2**–**6** in high purity.

3.3.3 Structure determination of TLN–2 to TLN–6 by X-ray crystallography

Crystal structures of **2**–**6** (chemical structures in Figure 3.1C) in complex with TLN were collected at high resolutions between 1.16–1.38 Å (Table 3.1). As shown in Figure 3.3, the conformations of the amino acids of their TLN binding sites are highly conserved and superimpose perfectly well, likewise the binding mode of the parent scaffold of all six ligands is virtually identical. The binding mode of this conserved part was already described extensively [33,93,135]. In brief, the scaffold coordinates monodentally the zinc ion with its negatively charged phosphonamidate oxygen. The carbamate group is disordered over two conformations with approximately equal occupancy. The *iso*-butyl portion of the ligand's leucine component is buried in the hydrophobic S_1' specificity cavity, a binding motif crucial for achieving high ligand affinity [33]. The S_1 pocket is occupied by a glycerol molecule from the cryobuffer, on top of which the carboxybenzyl moiety of the ligand is positioned. Also picked up from the buffer, a DMSO molecule is binding adjacent to the carbamate group of the ligand. Thus, **1**–**6** differ solely in their P_2' substituents and in the water networks adjacent to them. In the following, the positions of fifteen water molecules *W1*–*W15* (referring to water molecules found at distinct positions in the first solvation layer around the P_2' groups) are described according to the nomenclature and relative to the view angle chosen in Figure 3.4. At the right upper rim of the S_2' pocket, a second glycerol molecule is found (Figure 3.4A–F), which establishes weak H-bonds (distance >3.2 Å) to water molecules *W10* and *W11* in some of the crystal structures. We observed the two glycerol molecules in all 19 crystal structures that we determined in the previous studies [15,93,94,96]. A glycerol molecule is well known to replace three water molecules as a kind of rigidified surrogate in a crystal structure. Nonetheless, to validate and exclude whether these glycerol molecules take any artificial influence on the ligand pose and adjacent water structure, we succeeded to establish alternative crystallization conditions also yielding well-diffracting crystals using PEG400 and methyl-2,4-pentanediol (MPD) instead of glycerol as cryoprotectant. We also collected diffraction data of glycerol-free crystals of TLN in complex with **3**, **5** and **6** in order to validate whether the glycerol molecule exerts any artificial influence on the structural arrangement of the observed water molecules (Figure S3.2, Supporting Information). In summary, in the crystal structures with PEG400 and MPD as cryoprotectant, the positions of the glycerol OH groups are occupied by water molecules, and the established water networks in the S_2' pocket are very similar to the below described complexes with glycerol as cryoprotectant. The minor differences concern only highly mobile water molecules (borderline cases with respect to the placement of water

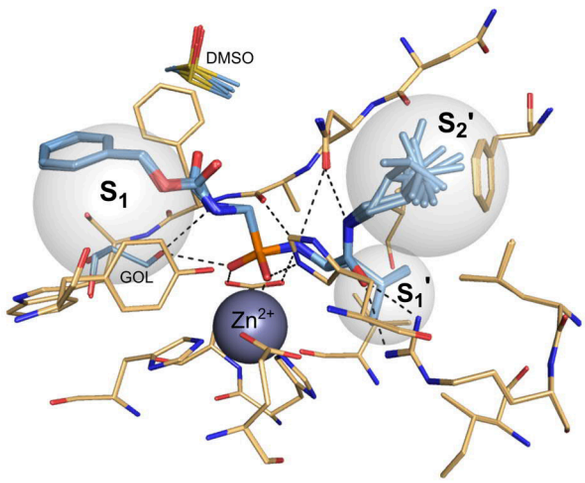


Figure 3.3. Superimposed crystal structures of the TLN–ligand complexes TLN-**1** to TLN-**6**. From all six crystal structures, the ligands, glycerol (GOL) and DMSO molecules are displayed with carbon atoms in light blue (color-coded hetero-atoms), TLN amino acids in beige, and the zinc ion as dark blue sphere. The S_1, S_1' and S_2' subpockets are indicated by semitransparent spheres. Polar interactions of at least moderate strength (≤3.2 Å) are indicated by black dotted lines. All crystal structures were superimposed on TLN-**1** by the alignment of amino acids within 5 Å of **1** (159 heavy atoms). The average RMSD is 0.05±0.02 Å as calculated by *fconv* [136]. As a result of the high rigidity of TLN, structural components of the binding site superimpose almost perfectly.

molecules in the refinement model) and the differences observed between the crystals exposed to different cryoprotectants fall into the same range as deviations recognized if diffraction data collected for different crystals of similar protein structures is compared. In the following, we compare the crystal structures in terms of observed electron densities in the S_2' pockets (F_o–F_c omit electron densities in Figure 3.4A–F) and refined *B* factors of the water molecules (*B* factors of all water molecules from the first solvation layer of TLN-**1** to TLN-**6** displayed as a heatmap in Figure 3.5) and we avoid to discuss only the presence or absence of a water molecule in the refinement model.

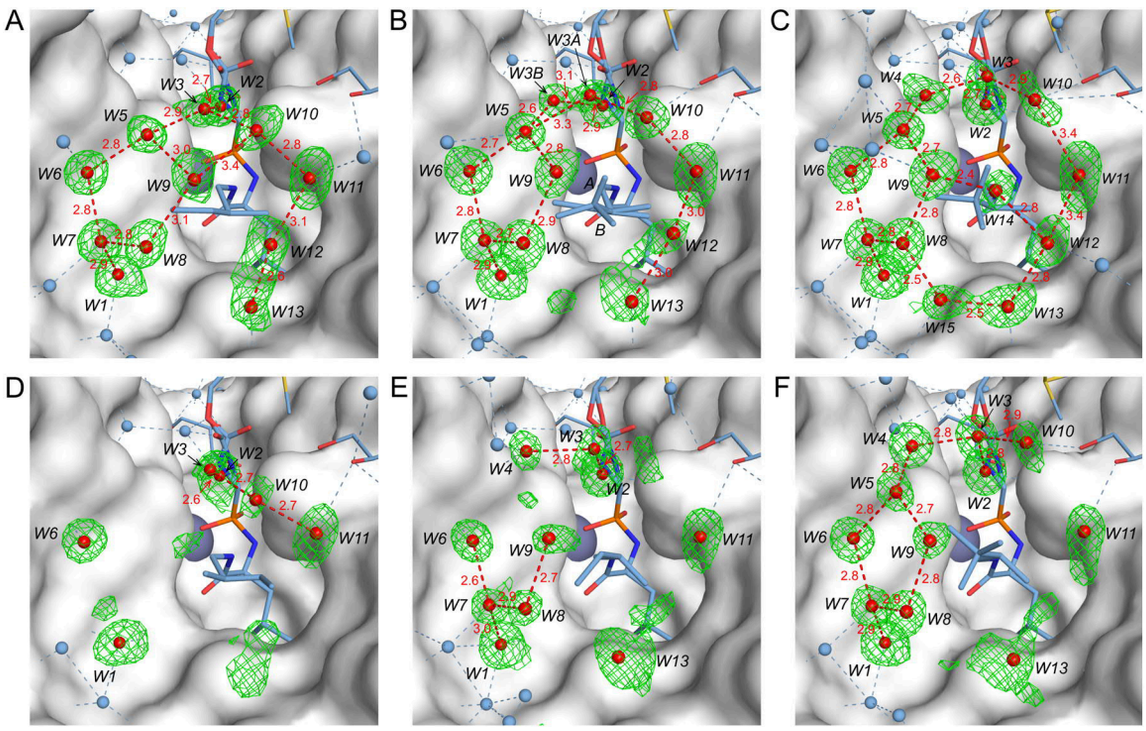


Figure 3.4. Crystal structures of TLN–ligand complexes **(A)** TLN-**1**, **(B)** TLN-**2**, **(C)** TLN-**3**, **(D)** TLN-**4**, **(E)** TLN-**5** and **(F)** TLN-**6**. In each panel, the solvent excluded surface of the respective crystal structure is displayed in white and the bound inhibitor is displayed as stick model (C blue, heteroatoms color-coded). The zinc ion of TLN is indicated as dark blue sphere partly buried by the surface. Water molecules from the first solvation layer of the ligands' P_2' groups are displayed as red spheres and labeled with an identifier (*W1–W15*) which is referred to in the main text; H-bonding distances between them are depicted as red dotted lines (maximum depicted distance: 3.4 Å, without hydrogens). The F_o–F_c omit electron densities of water molecules (green mesh) are displayed at a contour level of 3σ. For clarity reason, water molecules located outside of the first solvation layer of the P_2' groups are displayed as blue spheres with H-bonding distances indicated as blue dotted lines, and their electron densities are omitted. The two conformations of the P_2' group of **2** in complex with TLN (panel B) are labeled with *A* and *B*. F_o–F_c omit electron densities of the TLN-bound ligands are displayed in Figure S3.1 (Supporting Information).

Table 3.1. Data collection and refinement statistics for crystal structures TLN-**2** to TLN-**6** with glycerol as cryoprotectant.[a]

	Complex (PDB code)				
	TLN-**2** (5JT9)	TLN-**3** (5JS3)	TLN-**4** (5JXN)	TLN-**5** (5JVI)	TLN-**6** (5JSS)
(A) Data collection and processing					
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters: *a*, *b*, *c* (Å)	92.5, 92.5, 131.2	92.6, 92.6, 131.5	93.0, 93.0, 131.2	92.6, 92.6, 131.2	92.5, 92.5, 131.3
Matthews coefficient (Å^3/Da)[b]	2.4	2.4	2.4	2.4	2.4
Solvent content (%)[b]	48	48	48	48	48
(B) Diffraction data					
Resolution range (Å)	50–1.26 (1.34–1.26)	50–1.16 (1.23–1.16)	50–1.38 (1.46–1.38)	50–1.12 (1.19–1.12)	50–1.19 (1.26–1.19)
Unique reflections	89582 (14202)	114612 (18030)	69138 (10928)	127067 (20268)	106247 (16884)
$R(I)$sym (%)	8.7 (49.4)	6.6 (44.7)	5.3 (49.6)	7.4 (49.6)	9.5 (45.3)
Wilson *B* factor (Å^2)	8.4	8.6	11.9	8.3	8.3
Completeness (%)	99.9 (99.3)	99.5 (98.3)	99.9 (99.2)	100.0 (99.8)	100.0 (99.7)
Redundancy	15.0 (14.9)	16.1 (15.4)	9.7 (9.7)	19.4 (18.8)	15.1 (15.0)
$\langle I/\sigma(I)\rangle$	21.8 (5.5)	26.5 (5.8)	26.9 (4.3)	26.1 (6.0)	18.2 (5.2)
(C) Refinement					
Resolution range (Å)	34.20–1.26	40.10–1.16	46.51–1.38	30.36–1.12	40.1–1.19
Reflections used in refinement (work/free)	85102/4480	108881/5731	65681/3457	120713/6354	100934/5313
Final *R* value for all reflections (work/free) (%)	10.3/12.1	10.3/12.4	11.0/13.5	10.2/11.9	10.3/12.1
Protein residues	316	316	316	316	316
Calcium/zinc ions	4/1	4/1	4/1	4/1	4/1
Inhibitor atoms	30	31	29	30	31
Water molecules	424	428	415	435	424
RMSD from ideality:					
Bond lengths (Å)	0.009	0.01	0.011	0.011	0.011
Bond angles (°)	1.1	1.2	1.2	1.2	1.2
Ramachandran plot:[c]					
Residues in most favored regions (%)	88.1	88.9	88.1	87.8	88.1
Residues in additionally allowed regions (%)	10.7	10	10.7	11.1	10.7
Residues in generously allowed regions (%)	0.7	0.7	0.7	0.7	0.7
Residues in disallowed regions (%)[d]	0.4	0.4	0.4	0.4	0.4
Mean *B* factors (Å^2):[e]					
Protein non-hydrogen atoms	9.7	9.9	13.0	9.6	9.6
Protein Cα atoms	8.6	8.8	11.7	8.5	8.5
Inhibitor	9.1	9.6	16.1	9.8	10.1
Water molecules	25.8	26.0	30.0	25.5	25.2

[a] The crystal structure of TLN-**1** (PDB code 4MZN, resolution 1.17 Å) was already reported earlier [15]. Statistics of the glycerol-free crystal structures with PEG400 or MPD as cryoprotectant are listed in the Table S3.1 (Supporting Information). Numbers in parentheses represent the values of the highest resolution shells. [b] Matthews coefficients and solvent contents were calculated with the program *Matthews_coef* from the *CCP4* suite [137]. [c] Ramachandran plots were calculated with *PROCHECK* [138]. [d] The Ramachandran outlier is Thr26 as described in literature [139]. [e] Mean *B* factors were calculated with *MOLEMAN* [140].

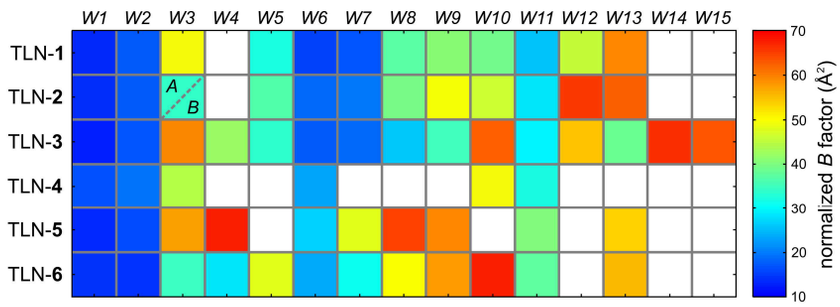


Figure 3.5. Heatmap of normalized *B* factors of water molecules *W1*–*W15* found in complexes TLN-**1** to TLN-**6**. Chemical structures of **1**–**6** are displayed in Figure 3.1C, water molecule numbering is according to Figure 3.4A–F. The normalized *B* factors are color-coded by a blue-to-red gradient. Individual *B* factor values of water molecules are omitted, since only distinct relative differences between *B* factors are discussed. A white field indicates no sufficient stabilization of the water molecule in the crystal structure for placement in the refinement model. *W3* in TLN-**2** is found in two conformations *A* and *B*. To compensate for differences between the *B* factor scales of different crystal structures as a result of varying crystal quality, *B* factors were normalized on the dataset with the lowest resolution (TLN-**4**, Table 3.1) and calculated as $B_{normalized} = B_{observed} \div B_{average(C\alpha)} \times B_{average(C\alpha)TLN\text{-}4}$, where $B_{observed}$ is the *B* factor of a water molecule as observed in the crystal structure, $B_{average(C\alpha)}$ is the average *B* factor of the Cα atoms of the respective crystal structure and $B_{average(C\alpha)TLN\text{-}4}$ is the average *B* factor of the Cα protein atoms in the crystal structure of TLN-**4**. No significant non-linearity is expected for the *B* factors of the datasets due to the narrow resolution range. Values for experimentally observed and normalized *B* factors are listed in Table S3.2 and Table S3.3 (Supporting Information).

3.3.4 Arrangement of water molecules in the S_2' pocket of TLN-1, TLN-2 and TLN-3

The structure of TLN-**1** was published previously [15] and served as starting point for the design of our new congeneric ligand series. Hence, the water networks observed in TLN-**2** to TLN-**6** are modifications of that in TLN-**1**. In TLN-**1**, the ligand places its 2(*S*)-methylbutyl P_2' group into the S_2' pocket (Figure 3.4A). The right cleft of the pocket is addressed by the terminal methyl group of the P_2' portion, whereas the terminal ethyl group is oriented toward the left. In TLN-**1**, *W5*–*W9* form a five-membered polygon with H-bond distances between 2.8–3.1 Å exhibiting low *B* factors (Figure 3.5). In total, twelve water molecules are detected in the first solvation layer around the P_2' substituent covering its large, solvent-exposed hydrophobic surface patch (Figure 3.6). Several of the water molecules are anchored to the protein surface via polar interactions: *W1* (Arg203 primary nitrogen, 3.0 Å; His231 carbonyl oxygen, 2.9 Å), *W2* (ligand **1** carboxybenzyl carbonyl O, 2.9 Å; Asn112 amide nitrogen, 3.0 Å), *W6* (Asp226 carboxy oxygen, 2.8 Å), *W11* (Asn111 backbone carbonyl oxygen, 2.8 Å) and *W13* (Tyr193, 3.1 Å). In contrast, *W3*, *W5*, *W7*, *W8*, *W9*, *W10* and *W12* are only stabilized

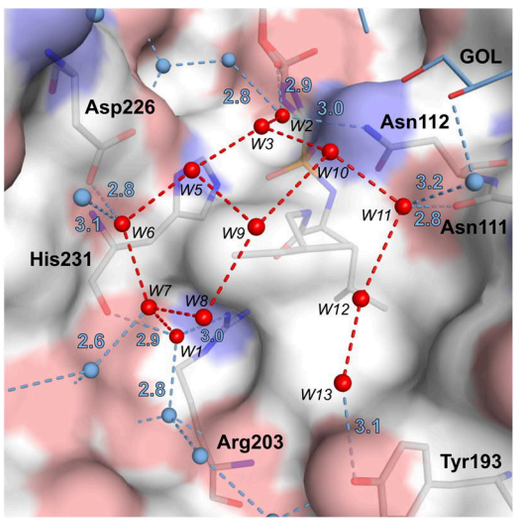


Figure 3.6. Topography of the S_2' pocket of TLN-**1**. The solvent excluded surface of the TLN–ligand complex is shown (color-coded, surface indicated for C in white, O in red, N in blue). Water molecules in the first solvation layer of the P_2' group are displayed as red spheres and labeled with identifiers according to Figure 3.4. Water molecules further remote than the first solvation layer and the glycerol molecule are displayed in blue. The position of **1** and TLN amino acids establishing H-bonds to water molecules of the first solvation layer are indicated by stick models below the surface. H-bonds between the water molecules of the first solvation layer are depicted as red dotted lines, H-bonds established to other water molecules, TLN amino acids or **1** are depicted as blue dotted lines (distances labeled in Å, maximum depicted distance: 3.4 Å).

by H-bonds to other water molecules or by van der Waals interactions with the apolar surface patch of the P_2' substituent of **1**.

In TLN-**2**, the 2(*S*),3-dimethylbutyl group of **2** orients its P_2' group similarly to **1**, and the additional methyl group is disordered over two positions (Figure 3.4B): conformation *A* (56% occupancy) is sticking out into the solvent (no contacts within hydrophobic interaction distance of ≤4.6 Å), whereas conformation *B* (44% occupancy) is oriented downward, alongside the protein surface. The observed disorder of the P_2' substituent is a result of the shallow, widely open S_2' pocket. Due to the steric requirement of the methyl group in conformation *A*, two distinct, mutually excluding sites are observed for *W3* (occupancies constrained to 50/50 in the refinement model). As a consequence of the reduced occupancy and owing to a strong correlation of *B* factors with occupancy [141], the refined *B* factor for this water molecule (Figure 3.5) has to be regarded with care and will hardly reflect its actual mobility. Furthermore, the distance between *W9* and *W10* increases from 3.4 Å in TLN-**1** to 4.5 Å in TLN-**2**, clearly exceeding the maximum distance for an energetically favorable H-bond [142,143]. The constraint which modifies the water structure between TLN-**1** and TLN-**2** is best visualized by use of the solvent accessible surface areas (SASAs; solvent excluded surface

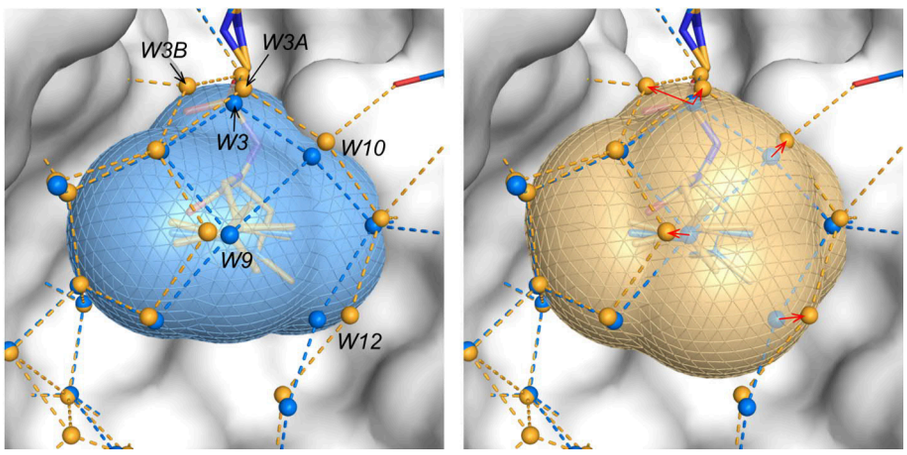


Figure 3.7. Differences in shape and size of the solvent accessible surfaces of TLN-**1** and TLN-**2**. In both images, the crystal structures TLN-**1** (blue) and TLN-**2** (orange) are superimposed. The semitransparent solvent accessible surface of the P_2' group of **1** is displayed in the left panel and of **2** in the right panel. In the right panel, shifts of water molecules observed between crystal structures TLN-**1** and TLN-**2** are indicated by red arrows. For TLN-**2**, both conformations *A* and *B* of the P_2' group of **2** are considered for the displayed solvent accessible surface. The solvent excluded surface of TLN is displayed in white.

area plus radius of a water molecule) as displayed for the P_2' portions in Figure 3.7: *W3*, *W9*, *W10,* and *W12* in TLN-**1** would penetrate inside the SASA of TLN-**2**, thus these water molecules in TLN-**2** must be shifted. Because of the expanded water structure, *W12* and *W13* become increasingly destabilized (Figure 3.5).

In TLN-**3**, the ligand exhibits a 2(*S*)-3,3-trimethylbutyl substituent (Figure 3.4C), representing the bulkiest P_2' portion in the (*S*)-configurated series. The binding mode of **3** closely resembles that of **2** if both conformations *A* and *B* of TLN-**2** would be merged. Two additional sites for water molecules (*W14* and *W15*) are refined, resulting in the formation of a six-membered polygon and well-defined electron density (Figure 3.4C) with strongly decreased *B* factor (Figure 3.5) of the incorporated *W13*. Furthermore, water molecules *W8* and *W9*, both participating in the five and six-membered polygonal water networks in TLN-**3**, are here better stabilized than in TLN-**2** (Figure 3.5). In TLN-**3**, *W10* is shifted distal from the P_2' substituent and is indicated by less electron density along with a higher *B* factor compared to TLN-**2**.

3.3.5 Arrangement of water molecules in the S_2' pocket of TLN-4, TLN-5, and TLN-6

Ligands **4**, **5** and **6** are the epimers of **1**, **2** and **3**. The overall quality of the crystal structure TLN-**4** is slightly lower compared to that of the other five complexes (resolution, R_{work}/R_{free} values, Wilson *B* factor; see Table 3.1). Nevertheless, only about ten water molecules less are observed in TLN-**4** at a total amount of >400. In TLN-**4** (Figure 3.4D), the P_2' portion is flipped over by 180° compared to TLN-**1**: The terminal P_2' ethyl group is oriented toward the right rim, whereas the P_2' methyl group is directed to the left of the S_2' pocket. Only six water molecules are detected in the crystal structure adjacent to the P_2' group mainly stabilized via H-bonds to protein residues.

In TLN-**5** (Figure 3.4E), in contrast to **4**, the P_2' substituent of **5** adopts a conformation analogous to that of the (*S*)-configurated P_2' substituents of **1** to **3**. One methyl group from the terminal *iso*-propyl portion of **5** is not defined in the electron density most likely owing to enhanced mobility. It was therefore not modelled in the structure. The absence of this methyl group is also observed in TLN-**5** with MPD as cryoprotectant (Figure S3.3B, Supporting Information). The missing of water molecules *W5*, *W10* and *W12* results in an incomplete water network in TLN-**5** (compared to that of the epimeric TLN-bound **2**). In the crystal structure TLN-**5** with MPD as cryoprotectant, *W5* is highly mobile but sufficient electron density is detected to allow placement of a water molecule in the refinement model (Figure S3.2B, Supporting Information). Furthermore, in the structure of TLN-**5** with glycerol as cryoprotectant, some F_o–F_c difference electron density is observed at the positions of *W5* and *W10* (Figure 3.4E), which, however, is too weak to justify placement of a fully populated water molecule in the refinement model of TLN-**5**. In TLN-**6** (Figure 3.4F), *W5* and *W10* were added to the refinement model, but they refined as highly mobile (Figure 3.5). Furthermore, *W10* is missing in the crystal structure of TLN-**6** with MPD as cryoprotectant (Figure S3.2C, Supporting Information). Consequently, the water molecules *W5* and *W10* are highly mobile in TLN-**5** and TLN-**6**, and the local concentration of their electron density is at the borderline for water placement in the refinement model. Thus, the water networks of TLN-**5** and TLN-**6** are highly similar. Overall, in the (*R*)-series a lower amount of water molecules with increased residual mobility (especially of *W5*–*W10*, Figure 3.5) is observed compared to the (*S*)-series.

3.3.6 Thermodynamic signatures of TLN-ligand complex formation measured by ITC

As we recently documented, ITC measurements comparable on the same scale and with minimal error margins can only be obtained if all ligands are studied with the same optimized

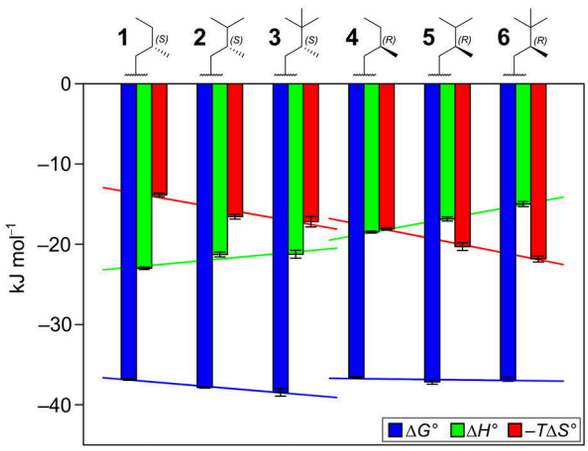


Figure 3.8. Thermodynamic parameters of **1**–**6** upon binding to TLN measured by ITC. The basic scaffold of the displayed P_2' groups is displayed in Figure 3.1A. The columns and the error bars represent the mean values calculated out of three measurements and their standard deviations, respectively. Trendlines of **1**→**3** (*S*) and **4**→**6** (*R*) for Δ*H*° (green), –*T*Δ*S*° (red) and Δ*G*° (blue) were calculated with *SigmaPlot*. Data values are listed in Table S3.4 (Supporting Information).

measurement protocol using the same protein batch [96]. We therefore reevaluated **1** along with **2**–**6** in the present study (for further details see Chapter 3.10.4, Supporting Information). Across the (*S*)-configurated series (**1**→**2**→**3**), the binding affinity Δ*G*° improved with growing number of methyl groups (Figure 3.8). Remarkably, this effect is determined by an increasingly favorable –*T*Δ*S*°, which is only partly compensated by a loss in Δ*H*° (slope of –*T*Δ*S*° is steeper than of Δ*H*°), leaving overall a gain in Δ*G*°. Interestingly, for the (*R*)-configurated series (**4**→**5**→**6**), no affinity enhancement is detected. The mutual compensation of Δ*H*° and –*T*Δ*S*° fully nullifies any affinity improvement as the Δ*H*° compensation is stronger (slopes of –*T*Δ*S*° and Δ*H*° are equal with opposite sign) compared to the (*S*)-series. Accordingly, in the (*S*)-series, a small but significant advantage in Δ*H*° is experienced parallel to the growing of the P_2' substituent into the S_2' pocket relative to the (*R*)-series.

To further validate whether the glycerol molecules found in our cryoprotected crystal structures have any distorting effect on the thermodynamic signature, we performed ITC titrations with the addition of different concentrations of glycerol (up to 10%, Figure S3.6 in the Supporting Information). These titrations revealed a systematical increase of Δ*H*° with increasing glycerol concentration paralleled by a compensating decrease of –*T*Δ*S*°. Most importantly, the relative difference between Δ*H*° and –*T*Δ*S*° remained unchanged, thus no specific effect and only an overall systematic influence of the added glycerol was observed. Comparable systematic influences, for example by the type of salt (NaCl or NaSCN) and its

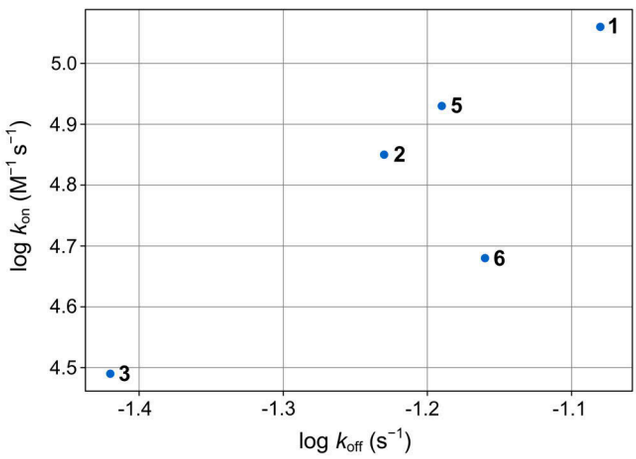


Figure 3.9. Kinetic map (log k_{on} vs. log k_{off}) of **1**–**3**, **5** and **6** as determined by global analysis of the single-cycle SPR measurements performed in triplicate for each ligand. Measurement results and SPR sensorgrams are shown in Table S3.7 and Figure S3.8 (Supporting Information). The results and standard deviations of the kinetic parameters as determined by individual analysis of the SPR sensorgrams are listed in Table S3.8 (Supporting Information).

concentration used in the measurement buffer, have been described previously [96]. Similarly, systematic influences on $\Delta H°$ and $-T\Delta S°$ were also observed for the measurement with the addition of different concentrations of DMSO (Figure S3.7, Supporting Information). These findings underscore that ITC data should only be recorded applying highly comparable measurement conditions and evaluated relative to each other in congeneric compound series [96].

3.3.7 Binding kinetics of TLN-ligand complex formation measured by SPR

The binding kinetic parameters of **1**–**3** and **5**–**6** (Figure 3.9) were determined by single-cycle SPR measurements preformed in triplicate for each ligand. Kinetic analysis of the SPR sensorgrams was performed by global analysis of the triplicate data applying a 1:1 binding model [144,145], which agreed well with individual analysis of the sensorgrams. The relative standard deviations of the individual analysis results were about 30% for k_{on} and 10% for k_{off} (Table S3.8, Supporting Information). Ligand **4** was not tested, as this ligand was synthesized at a later stage of the study. As depicted on the kinetic map (Figure 3.9), **1** showed the fastest k_{off} within the series, whereas the dissociation rates for **2**, **5** and **6** are slower and fall within a narrow window. Ligand **3** instead shows a significantly prolonged dissociation rate compared to all other members of the series.

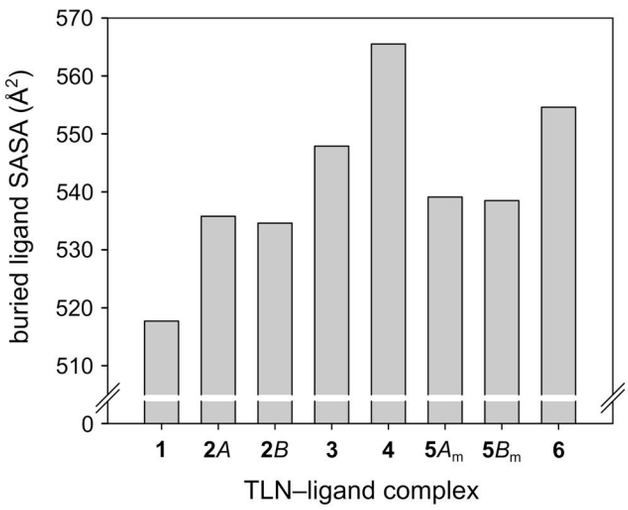


Figure 3.10. Buried solvent accessible surface areas (SASAs) of **1**–**6** while in complex with TLN. The P_2' group of **2** adopts two conformations in the crystal structure (*A* and *B*, Figure 3.4). One methyl group of the P_2' portion of **5** is not visible in the electron density and was thus modeled in two conformations **5**A_m and **5**B_m (Figure S3.9, Supporting Information). Calculated data values are listed in Table S3.9 (Supporting Information).

3.3.8 Buried SASAs of the TLN-ligand complexes

Figure 3.10 shows the computed buried solvent accessible surface areas (SASAs) of **1**–**6** in complex with TLN. One methyl group of the P_2' portion of **5** remained undetected in the electron density (Figure 3.4E). Therefore, the missing methyl group was modeled in two conformations A_m and B_m based on the crystal structure (Figure S3.9, Supporting Information) and the buried SASAs of these two conformations were calculated. The SASA buried within the (*S*)-series increases monotonously from **1**→**2**→**3** by approximately 15 Å^2 per added methyl group. TLN-**4** exhibits the largest buried SASA of all six ligands and the buried SASAs of **5** and **6** are slightly larger than those of their respective epimers **2** and **3**.

3.4 Discussion

In the current study, we wanted to validate our hypothesis that ligand binding to an open, rather flat and solvent-exposed binding pocket can be enhanced by optimizing the surface water network wrapping around exposed parts of the bound ligand. We started with the

previously characterized peptidomimetic TLN inhibitor **1** [15], and modified its hydrophobic solvent-exposed P_2' substituent attached to the parent scaffold. We improved binding affinity by maximizing the desolvation of the increasingly bulky P_2' substituents along with an enhancement of the water network wetting the surface of the formed complex. Moreover, we expected that the residence time of the complex could be expanded with increasing quality and perfection of the formed water network. Prior to synthesis, we predicted the putative water pattern around the designed ligands by MD simulations.

In our previous purely descriptive studies, we observed that small changes of the solvent-exposed ligand surface can strongly modulate the stability and complexity of the formed water network adjacent to the bound ligands [15,93,94]. Since the spatial arrangement of water molecules across apolar surfaces is governed by a complex architecture and the addition of a sole methyl group can already lead to the unfavorable disruption of the adjacent water network, optimal hydration of a partly solvent-exposed apolar P_2' substituent is a challenging design task.

As starting point we chose **1**, which already displayed rather potent inhibition properties but showed local deficiencies in the solvation pattern next to its P_2' substituent. This was indicated by a comparison with structures of closely related complexes. By merging features of their P_2' substituents with those of **1**, we designed a small series of ligands comprising chiral aliphatic P_2' substituents. To predict their impact on the quality of the wetting surface water network, we followed our recently introduced MD approach [133]. To further validate our hypothesis, we considered both chiral orientations, as the simulations suggested significant differences between the stereoisomers. Since the designed ligands exhibited a second stereogenic center at P_1', epimeric pairs of ligands resulted. This leads to the disadvantage that differences in the desolvation cost to transfer the corresponding ligands of the epimeric pairs from the bulk solvent phase to the protein pocket cannot be entirely excluded. However, as the two stereogenic centers are separated by several bonds, we assume quite similar physicochemical properties for the matching epimeric pairs. Across the series, the MD analysis suggested small but significant differences in the completeness of the surface water network formed next to the different P_2' substituents rendering the (*S*)-configurated **3** as most promising candidate of the series. Subsequently, we synthesized the ligands stereoselectively and characterized the complexes formed via crystallography, microcalorimetry and surface plasmon resonance. Concerning the crystal structure analysis, in all cases we obtained diffraction data with very high resolution, also falling into a narrow window (mean: 1.22±0.10 Å,). This is important to reliably compare details of the water structures between the different crystal structures, as deviating resolution can affect such details and will complicate the mutual matching of *B* factors [146].

The parent scaffold of all six ligands adopts virtually the same binding pose (Figure 3.3). Thus, the observed differences between the studied ligands **1**–**6** most likely originate predominantly from the desolvation differences of the gradually increasing and partly buried P_2' substituents and from deviations of the formed surface water networks "wetting" the newly formed complexes. They show varying degrees of completeness and perfection, which in turn, influence the thermodynamic and binding kinetic signature of complex formation. Whereas the fixation of water molecules on the surface of the bound ligand increases binding enthalpy and reduces entropy, in contrast binding entropy is favored and enthalpy lowered by enhanced mobility up to the displacement of water molecules into the bulk water phase. The mobility and occupancy of individual water molecules is indicated by the spatial concentration of the electron densities (Figure 3.4) and the assigned *B* factors (Figure 3.5). A strong fixation of a water molecule results from the embedding into a geometrically rather constrained H-bonding network, also involving the formation of H-bonds to adjacent functional groups of the protein's amino acids.

Although essential in the current series, the arrangement of water molecules is possibly only one component determining the thermodynamic signature of the binding process. The total thermodynamic signature is in any case the sum of many contributions and might be composed of partly compensating or mutually enhancing contributions. Hence, it is even more important that we attempt to only evaluate relative differences of the studied complexes and not their absolute values. Furthermore, ligand binding can be accompanied by global conformational adjustments of the protein partly masking the thermodynamic signature of the local binding event. However, from our experience with the system, global adjustments of TLN are unlikely. The enzyme has proven to be highly rigid and the sole differences introduced between the congeneric ligands are their solvent-exposed P_2' groups. Fenley et al. recently evaluated a large-scale MD trajectory of BPTI (bovine pancreatic trypsin inhibitor) and observed remarkable transitions between states of unchanged overall Gibbs free energy but significantly altered enthalpy/entropy inventory [147]. This entropy–enthalpy transduction might suggest a physical mechanism underlying entropy–enthalpy compensation in such systems. However, we propose that in our congeneric ligand series, where binding occurs to a rigid protein, the ligands always address the same or very similar configurations of TLN.

3.4.1 Prediction of solvation sites by MD simulations and their agreement with crystallographically determined solvation sites

Based on our earlier study on four TLN ligands [94], we developed an MD simulation protocol to predict water networks adjacent to solvent-exposed ligand groups [133]. This protocol correctly reproduced the rupture of a water network between two ligands differing by only one

single methyl group [94]. The rupture was responded by a dramatic loss in binding enthalpy ($\Delta\Delta H°_{methyl\rightarrow H}$ = +13.3±0.6 kJ mol^{-1}) and an increase in binding entropy ($-T\Delta\Delta S°_{methyl\rightarrow H}$ = –7.7±0.4 kJ mol^{-1}), overall resulting in a lowered affinity ($\Delta\Delta G°_{methyl\rightarrow H}$ = +5.7±0.3 kJ mol^{-1}). This example underlines the pronounced effect of the ligand-capping water network on the thermodynamic signature.

Importantly, our MD simulation approach does not require any *a priori* knowledge about water positions but allows the prediction of hydration sites adjacent to protein–ligand complexes from scratch in agreement with experiment. In the present study, the tool was applied to predict TLN-**2,** TLN-**3**, TLN-**5** and TLN-**6**. Its predictive power could be further assessed by simulating TLN-**1**, as its crystal structure had been determined prior to the present study [15]. For TLN-**1**, the computed hydration sites match remarkably well with the difference electron density assigned to water molecules by crystallography (Figure 3.2B). Only the population of *W8* is underestimated, which is in line with our previous observation that water-to-methyl interactions are predicted as too weak by the *AMBER* force field [133].

To estimate whether the designed P_2' substituents exhibit higher or lower hydration propensities than those of **1**, the predicted solvation sites were mutually compared. The simulations of TLN-**1** and TLN-**3** suggest some advantages of the latter adjacent to the P_2' substituents (Figure 3.2B,C). This should render **3** superior to **1** with respect to affinity, since the enthalpic component of stabilizing the water network is larger and can compensate for the increasing enthalpic cost to (entropical beneficially) desolvate **3** over **1**, as its substituent comprises two additional methyl groups, resulting in a significantly larger buried surface area (Figure 3.10). For TLN-**6**, a possible rupture of the water network in the center of the apolar surface next to the P_2' substituent is suggested. Facing the subsequently determined crystal structures with our predictions, TLN-**3** (Figure 3.11A) exhibits the water molecules capping the apolar P_2' substituent which are likewise predicted too weak (as *W8* in TLN-**1**). The in the crystal structure well-stabilized *W8* and *W9* are weakly indicated by the MD approach, and the relatively mobile *W14* (high *B* factor and weak electron density) is not predicted by the computer analysis, at least on the 48% threshold level. *W15* is correctly predicted by a tubular solvation site, which also hosts *W13*. In TLN-**6**, *W8* and *W9* are not predicted by MD, which, however, agrees with the experimental observation that these water molecules are significantly less stabilized in TLN-**6** than in TLN-**1** or TLN-**3** (cf. high *B* factors, Figure 3.5). The shifted position of *W10* was correctly predicted, resulting in a large gap toward *W11* beyond H-bonding distance. Water molecule *W12* is not observed in the electron density, whereas the MD simulation predicts a hydration site at this position. In summary, TLN-**1**, TLN-**3** and TLN-**6** are convincingly predicted on qualitative level and the relative ranking of the epimer complexes TLN-**3** (strong fixation) and TLN-**6** (weak fixation) was correctly assigned.

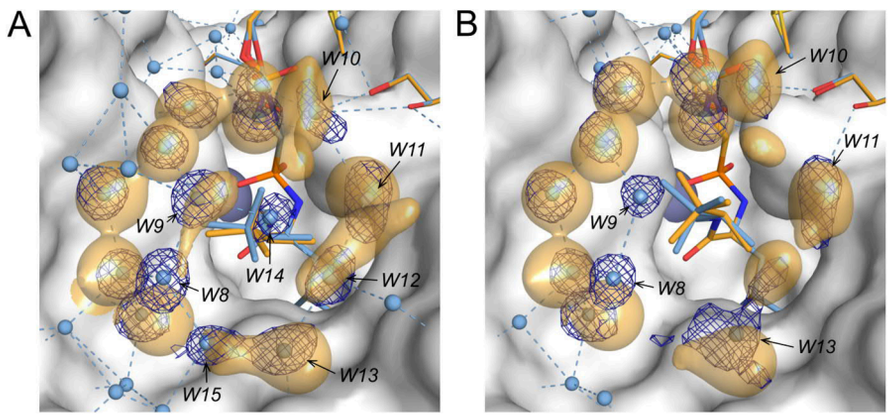


Figure 3.11. Prediction of water solvation sites as calculated by MD simulations in comparison to the crystallographically observed electron densities and refined crystal structure models of **(A)** TLN-**3** and **(B)** TLN-**6**. The modeled coordinates of ligand, glycerol and DMSO molecules used in the MD simulations are displayed as yellow stick models with color-coded heteroatoms. The yellow, semitransparently contoured regions show computed areas in the first solvation layer of the P_2' substituents with an occupancy probability by a water molecule of at least 48%. The crystallographically determined binding modes of ligand and additive molecules are superimposed as blue stick models with color-coded heteroatoms. Water molecule positions determined in the crystal structures are displayed as blue spheres, and the F_o–F_c omit electron density is displayed as dark blue mesh at a contour level of 3σ for the water molecules positioned in the first solvation layer of the P_2' groups. H-bond distances are indicated as blue dotted lines. Positions of water molecules, which are discussed in the main text, are labeled with identifiers according to Figure 3.4. The solvent excluded surface of TLN is shown in white.

The predicted solvation sites in TLN-**2** and TLN-**5** differ more strongly from the crystallographically observed electron densities. This results from the disorder of the P_2' substituents of **2** and **5** indicated in the crystal structures, which was difficult to predict as the disorder was not considered in the MD simulations (Figure S3.10, Supporting Information).

3.4.2 Correlation of structural data with thermodynamic signature of complex formation

Our starting ligand **1** has been established as most potent binder from the previously studied series (Figure 3.1B) [15]. We related its superior affinity to an entropically beneficial burial of its surface along with the establishment of an extensive enthalpically favored surface water network. In TLN-**1**, a particularly favorable five-membered water polygon is formed (Figure 3.4A, *W5*–*W9*), which is well-known for its favorable energetic contribution [148,149].

Compared to TLN-**1**, the additional methyl group in TLN-**2** creates disorder over two conformational states, perturbing the neighboring water structure (Figure 3.7) and increasing

the mobility of the adjacent water molecules: *W3* is observed in two orientations, and *W9*, *W10*, *W12* and *W13* show enhanced *B* factors (Figure 3.5). Consequently, enthalpy decreases from TLN-**1** to TLN-**2** (Figure 3.8). This is overcompensated by a more favorable entropy resulting from the burial of a larger apolar surface area (Figure 3.10) and the enhanced sidechain mobility, which is entropically beneficial. Altogether, the more favorable entropy contribution is only partly compensated by enthalpic losses, resulting overall in a slightly enhanced affinity of **2** over **1** by $\Delta\Delta G°_{1\to2} = -1.0\pm0.1$ kJ mol^{-1}.

The thermodynamic signature of **3** is comparable to that of **2**, with a slightly elevated entropic and virtually unchanged enthalpic term leading to an increased affinity of **3** over **2** (Figure 3.8). The surface water networks of both complexes differ in several regards. In TLN-**3** (Figure 3.4C), a complete six-membered water network polygon (*W8*, *W9*, *W12*–*W15*) is established adjacent to the five-membered one and is integrated in an eight-membered ring structure (*W3*–*W5*, *W9*–*W12*, *W14*). Such fused polygonal water arrangements can be considered as optimal solvation shell to coat the surface of a formed protein–ligand complex in terms of H-bonding and thus inherent enthalpy contributions. Overall, TLN-**3** shows the most perfect water network along with an increased apolar surface burial compared to TLN-**2**, resulting in a slightly superior affinity. This leads to an unchanged enthalpic contribution as desolvation costs are compensated by the formed enthalpically favored water structure. For **3**, however, an entropic benefit remains originating from the burial and desolvation of the additional methyl group ("classical" hydrophobic effect).

The complexes TLN-**5** and TLN-**6** formed with (*R*)-configurated ligands also show elaborate surface water networks involving the formation of the stabilizing five-membered water polygon (*W5*–*W9*). Detailed analysis suggests that the involved water molecules experience much higher *B* factors than in the corresponding complexes with the (*S*)-configurated epimers (Figure 3.5). This supposedly less stable arrangement results from the inverted stereochemistry and increases the steric demand of the (*R*)-configurated P_2' substituents. Furthermore, one terminal methyl group of **5** is not detectable in the crystal structure suggesting enhanced mobility, likely increasing entropy and reducing the stabilization of the adjacent water network. In contrast to the (*S*)-series, the (*R*)-configurated substituents do not enhance affinity (Figure 3.8). The gradual enhancement in $-T\Delta S°$ with growing number of methyl groups is similar within the (*S*) and (*R*)-series. However, in the (*R*)-series the loss in $\Delta H°$ completely nullifies the advantage in $-T\Delta S°$. Thus, overall the affinity enhancement across the (*S*) series ligands results from an enthalpic advantage of the growing P_2' substituents. They achieve more elaborated and energetically improved surface water networks.

Ligand **4** shows an unexpected binding mode, as its P_2' group is flipped by 180° compared to the other five ligands (Figure 3.4D). Remarkably, this ligand shows the largest surface burial across the series, significantly higher than that of its epimer **1** (Figure 3.10). The flipped

orientation takes considerable impact on the established water structure, showing a much lower amount of recruited water molecules compared to the other TLN complexes. This results in the striking observation that TLN-**4**, even though exhibiting the largest surface burial, shows the lowest affinity across all six ligands (Figure 3.8). This underpins our observation that the sole burial of hydrophobic surface portions of a ligand is clearly not sufficient to explain the binding features. Due to the considerable conformational change of the P_2' substituent, a direct comparison of $\Delta H°$ and $-T\Delta S°$ of TLN-**4** with the thermodynamic signatures of the other five ligands is complicated.

3.4.3 Kinetic analysis of the ligands

The overall rather slow association of ligands binding to TLN is likely governed by a large conformational transition (induced fit) of the protein, rendering influences of individual ligands on k_{on} rather insignificant. As a matter of fact, the experimental determination of association rate constants is dependent on the concentration of the studied samples and thus prone to additional experimental uncertainties (e.g. weighting errors or repeated freeze-thaw cycles of the inhibitor solution affecting its concentration). For these two reasons, we refrain from a detailed interpretation of k_{on}. This uncertainty also affects K_d values determined by SPR, and will also afflict a direct comparison with K_d values from ITC measurements. In any case, there might be inherent differences from a theoretical point of view between both techniques involved, making a direct comparison of K_d values taken from both methods difficult. ITC observes a system under thermodynamic equilibrium conditions based on a particular binding model, whereas SPR records a steady state situation in a flow cell using immobilized protein, which does not necessarily relate to the same binding model as different structural states might determine binding kinetics. This may lead to differences in the determined K_d values [150].

Based on the thermodynamic equilibrium K_d values determined by ITC, **3** clearly shows the highest affinity of the series (Figure 3.8), whereas in the SPR measurements the second largest K_d value was determined for this inhibitor (Table S3.7, Supporting Information). Ligand **3** shows, however, a significantly longer residence time compared to all other ligands (Figure 3.9). Since the local interactions of a specific inhibitor conceivably have a higher influence on the dissociation kinetics, we relate this decreased dissociation rate constant of **3** to the formation of the pronounced surface water network caging the hydrophobic P_2' substituent to stabilize the complex and thus prolong residence time of the ligand.

3.5 Conclusion

Drug optimization aims for the tailored optimization of binding parameters to endow a ligand with the required potency, selectivity and binding efficacy. We increasingly recognize that the sole optimization to enhanced binding affinity is not sufficient to render a ligand as most promising candidate for further development. Consequently, additional parameters such as thermodynamic and binding kinetic signatures are consulted to obtain a more detailed view on the binding process. Furthermore, increasing resolution of the crystal structures determined across narrow series of protein–ligand complexes discloses tiny differences in the binding poses, and adaptations of the target protein or modulations of the "wetting" surface water networks. For the medicinal chemist who performs ligand optimization by means of chemical synthesis, it is essential that these modulations, which finally improve the ligand's profile, result from properties of the bound ligand and its partly solvent-exposed substituents. They allow fine-tuning of affinity, enthalpy, entropy and binding kinetic properties, as they are accomplished by well-established medicinal chemistry optimization steps. In the present study, we show that by means of optimizing the composition of a partly exposed, apolar ligand substituent bound to a flat, solvent-exposed binding pocket of a protein, the relevant binding parameters can be fine-tuned using rational design principles. We therefore had to analyze, predict and characterize the substituent's burial and in parallel the quality and perfection of the adjacent formed surface water network, which coats the formed complex. The advantage of this concept is that the chemical adjustments needed to drive the thermodynamic and kinetic parameters into a desired range are performed using the normal toolbox available to medicinal chemists. Our strategy requires the following steps along an iterative design cycle: *(i)* molecular design of the exposed substituent to optimize the pocket burial and adjacent surface water layer, *(ii)* molecular dynamics simulations to validate the proposed surface water network, *(iii)* ligand synthesis, followed by *(iv)* structural, microcalorimetric and binding kinetic characterization of the formed complexes. The Table of Contents Graphic (Figure 8.3) shows the stepwise affinity enhancement across our studied series. The most potent complex TLN-**3** (with the 2(*S*)-3,3-trimethylbutyl P_2' substituent, at the far right of the diagram in the Table of Contents graphic) is by about 1.5 orders of magnitude more potent in terms of affinity than the initial purely methylated complex (at the far left of the diagram). As the detailed thermodynamic characterization shows, this only partly results from enhancements of the "classical" hydrophobic effect. Moreover, additionally important, mainly more enthalpy-driven effects result from the optimization of the surface water network that coats TLN-**3** almost perfectly and establishes several fused polygonal water arrangements, which are characterized by a particular stability. Apart from the enhancement of the thermodynamic profile, TLN-**3** shows prolonged residence time that results from a more stable protein–ligand complex. Obviously, the optimized surface water layer captures and holds the ligand more tightly to the protein, thus increasing the barrier for its release. It is possible that this is a rather

general concept to modulate binding kinetics, namely by enhancing the interaction of a bound ligand with the adjacent surface water network. Exploitation of this property might allow the medicinal chemist to fine-tune binding kinetic parameters via ligand optimization for many drug targets.

3.6 Experimental Section

3.6.1 Water Network Prediction by Molecular Dynamics Simulations

The crystal structure of TLN-**1** (PDB code 4MZN) was used for modeling of the ligands as well as for the MD simulations. In order to provide a common environment for modeling and simulation, the Cartesian coordinates of protein, zinc ion, and cryobuffer molecules (DMSO, glycerol) were used. The preparation was performed similarly to our protocol described previously [133]. After protonation, all crystallographically observed water molecules were removed. All ligands were modeled within the binding site of TLN-**1**. As template structures for the modelling of the ligands, **1** was used for the (*S*)-configurated **2** and **3**, and the ligand in its complex to TLN from the PDB entry 4MTW [15] was used as a template structure for the (*R*)-configurated **5** and **6**, as they provided suitable exit vectors. Modeling and a subsequent minimization of the S_2' groups was performed using the molecule builder function and the *AMBER99* force field implemented in *MOE* [151]. Atomic charges for the ligands were calculated with the *RESP* methodology [152] based on quantum mechanical calculations obtained by *Gaussian09* [153] at HF/6-31G* level. The MD simulations were performed with the *AMBER14* package [154], using the ff99SB force field and periodic boundary conditions. During all simulation steps, all atoms, except hydrogen atoms and water molecules, were restrained to their coordinates of the crystal structure. In a 20 ns production phase, water molecule positions were recorded every 2 ps. This trajectory was analyzed to calculate the solvation sites using the *VOLMAP* plugin in *VMD* [155]. The protocol is described in detail in our earlier contribution [133].

3.6.2 Ligand Synthesis and Purification

^{1}H, ^{13}C and ^{31}P NMR spectra were recorded on a JEOL ECX-400 or JEOL ECA-500 instrument. All chemical shift values are reported in ppm relative to the non-deuterated solvent signal. An external standard was used for ^{31}P NMR spectra (referenced to: 85% H_3PO_4) and ^{13}C NMR spectra in D_2O (referenced to: trimethylsilyl propanoic acid). ESI-MS spectra were recorded on a Q-Trap 2000 system by Applied Biosystems. For the description of multiplicity the following abbreviations were used: s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, dd = doublet of doublet, m = multiplet, br = broad signal. For high resolution ESI-MS

a LTQ-FT Ultra mass spectrometer (Thermo Fischer Scientific) was used. EI-MS analysis was carried out on a Micromass AutoSpec instrument. For HPLC chromatography a Shimadzu LC-20 system equipped with a diode array detector was used. Analytic separations were carried out with a MN Nucleodur 100-5 C18 ec 4.6×250 mm column using a water-acetonitrile gradient. For semi-preparative separations a Water XSelect CSH C18 10×250 mm column using a water-acetonitrile gradient was used. The purity of all inhibitors used for biophysical experiments was greater than 95%, as determined by HPLC

General procedure for the synthesis of acyloxazolidinones **9**–**13***: n*-BuLi (1.2 eq) was slowly added to a solution of **7** or **8** (1.0 eq) in THF at -78 °C under argon. The solution was allowed to warm to room temperature over 30 minutes. The respective acid chloride (1.1 eq) was added to the yellow solution and the reaction mixture was stirred for 2 h at -78 °C. The reaction was quenched with saturated NH_4Cl solution and extracted with EtOAc (3×20 mL). The combined organic extracts were washed with brine and dried over $MgSO_4$. The crude product was purified by silica gel column chromatography (cyclohexane/EtOAc 5:1).

(S)-4-Benzyl-3-(3-methylbutanoyl)oxazolidin-2-one (**9**): Compound **9** was synthesized according to the general procedure using *n*-BuLi (2 M in hexanes, 2.4 mL, 6.00 mmol), *(S)*-4-benzyloxazolidin-2-one (**7**, 886 mg, 5.00 mmol) and 3-methylbutanoyl chloride (663 mg, 5.50 mmol). The product was obtained as a colorless oil (1243 mg, 4.76 mmol, 95%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.02 (t, *J* = 6.6, 6H), 2.17–2.28 (m, 1H), 2.71–2.76 (m, 1H), 2.79 (dd, *J* = 10.2, 6.0, 1H), 2.89 (dd, *J* = 16.2, 6.7, 1H), 3.31 (dd, *J* = 13.3, 3.3, 1H), 4.12–4.23 (m, 2H), 4.63–4.74 (m, 1H), 7.18–7.39 (m, 5H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 22.6, 22. 7, 25.1, 38.1, 44.1, 55.3, 66.2, 127.4, 129.1, 129.5, 135. 5, 153.6, 172.8. **MS (ESI+)** *m/z* calculated for $C_{15}H_{20}NO_3$ $[M+H]^+$: 262.32; found: 262.09.

(S)-4-Benzyl-3-(3,3-dimethylbutanoyl)oxazolidin-2-one (**10**): Compound **10** was synthesized according to the general procedure using *n*-BuLi (2.5 M in hexanes, 3.4 mL, 8.40 mmol), *(S)*-4-benzyloxazolidin-2-one (**7**, 1240 mg, 7.00 mmol) and 3,3-dimethylbutanoyl chloride (1036 mg, 7.70 mmol). The product was obtained as a colorless oil (1776 mg, 6.45 mmol, 92%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.09 (s, 9H), 2.71 (dd, *J* = 13.2, 10.1, 1H), 2.86 (d, *J* = 14.9, 1H), 2.99 (d, *J* = 14.9, 1H), 3.35 (dd, *J* = 13.3, 3.2, 1H), 4.07–4.22 (m, 2H), 4.62–4.78 (m, 1H), 7.16–7.43 (m, 5H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 29.7, 31.6, 38.2, 46.3, 55.5, 65.9, 127.4, 129.1, 129.5, 135.6, 153.6, 172.0. **MS (ESI+)** *m/z* calculated for $C_{16}H_{22}NO_3$ $[M+H]^+$: 276.35; found: 276.08.

(R)-4-Benzyl-3-butyryloxazolidin-2-one (**11**): Compound **11** was synthesized according to the general procedure using *n*-BuLi (2.5 M in hexanes, 6.6 mL, 16.50 mmol), *(R)*-4-benzyloxazolidin-2-one (**8**, 2660 mg, 15.00 mmol) and butyryl chloride (2238 mg, 21.00 mmol). The product was obtained as a colorless oil (3483 mg, 14.08 mmol, 94%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.01 (t, *J* = 7.4 Hz, 3H), 1.66–1.80 (m, 2H), 2.77 (dd, *J* = 13.3, 9.7 Hz, 1H),

2.83–3.01 (m, 2H), 3.30 (dd, *J* = 13.3, 3.2 Hz, 1H), 4.14–4.24 (m, 2H), 4.63–4.72 (m, 1H), 7.18–7.36 (m, 5H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 13.8, 17.8, 37.5, 38.1, 55.3, 66.3, 127.5, 129.1, 129.6, 135.4, 153.6, 173.4. **MS (ESI+)** *m/z* calculated for $C_{14}H_{18}NO_3$ $[M+NH_4]^+$: 265.33; found: 265.19.

(R)-4-Benzyl-3-(3-methylbutanoyl)oxazolidin-2-one (**12**): Compound **12** was synthesized according to the general procedure using *n*-BuLi (2.5 M in hexanes, 2.8 mL, 7.00 mmol), *(R)*-4-benzyloxazolidin-2-one (**8**, 1050 mg, 5.90 mmol) and 3-methylbutanoyl chloride (787 mg, 6.50 mmol). The product was obtained as a colorless oil (1368 mg, 5.23 mmol, 88%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 1.01 (dd, *J* = 8.3, 6.7, 2H), 2.15–2.29 (m, 1H), 2.71–2.81 (m, 2H), 2.89 (dd, *J* = 16.2, 6.7, 1H), 3.31 (dd, *J* = 13.3, 3.3, 1H), 4.09–4.23 (m, 2H), 4.63–4.73 (m, 1H), 7.18–7.37 (m, 5H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 22.5, 22.7, 25.1, 38.1, 44.1, 55.3, 66.2, 127.4, 129.0, 129.5, 135.4, 153.5, 172.8. **MS (EI)** *m/z* calculated for $C_{15}H_{19}NO_3$ $[M]^+$: 261.32; found: 261.

(R)-4-Benzyl-3-(3,3-dimethylbutanoyl)oxazolidin-2-one (**13**): Compound **13** was synthesized according to the general procedure using *n*-BuLi (2.5 M in hexanes, 3.4 mL, 8.40 mmol), *(R)*-4-benzyloxazolidin-2-one (**8**, 1240 mg, 7.00 mmol) and 3,3-dimethylbutanoyl chloride (1036 mg, 7.70 mmol). The product was obtained as a colorless oil (1876 mg, 6.81 mmol, 97%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 1.09 (s, 9H), 2.71 (dd, *J* = 13.3, 10.0 Hz, 1H), 2.86 (d, *J* = 14.9 Hz, 1H), 2.99 (d, *J* = 14.9, 1H), 3.34 (dd, *J* = 13.3, 3.3 Hz, 1H), 4.10–4.18 (m, 2H), 4.64–4.73 (m, 1H), 7.21–7.36 (m, 5H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 29.7, 31.5, 38.1, 46.2, 55.5, 65.9, 127.4, 129.0, 129.5, 135.6, 153.6, 172.0. **(MS ESI+)** *m/z* calculated for $C_{17}H_{24}NO_3$ $[M+H]^+$: 290.38; found: 290.13.

General procedure for the synthesis of acyloxazolidinones **14**–**18***:* A solution of diisopropylamine (1.3 eq) in dry THF under argon was cooled to -78 °C. n-BuLi (1.2 eq) was slowly added to the solution and the reaction mixture was allowed to warm to room temperature over 60 minutes. After cooling to -80 °C the respective oxazolidinone **9**–**13** (1.0 eq) was added dropwise to the mixture. After 60 minutes MeI (4.0 eq) was added to the solution. The reaction mixture was stirred for 5 h without further cooling. The reaction was quenched with saturated NH_4Cl-solution and extracted with EtOAc (3×20 mL). The combined organic extracts were washed with brine and dried over $MgSO_4$. The crude reaction product was purified by silica gel column chromatography (cyclohexane/EtOAc 6:1).

(S)-4-Benzyl-3-((S)-2,3-dimethylbutanoyl)oxazolidin-2-one (**14**): Compound **14** was synthesized according to the general procedure using diisopropylamine (591 mg, 5.84 mmol), n-BuLi (2.5 M in hexanes, 2.2 mL, 5.39 mmol), oxazolidinone **9** (1173 mg, 4.49 mmol) and MeI (2550 mg, 17.96 mmol). Recrystallization of the chromatographically pure product from cyclohexane gave the diastereomerically enriched product as a colorless solid (*dr* 20:1, 892 mg, 3.24 mmol, 68%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.90–0.99 (m, 6H), 1.17 (d, *J* = 6.4 Hz, 3H),

1.94–2.06 (m, 1H), 2.76 (dd, *J* = 13.3, 9.7 Hz, 1H), 3.29 (dd, *J* = 13.3, 3.1 Hz, 1H), 3.54–3.67 (m, 1H), 4.12–4.23 (m, 2H), 4.63–4.74 (m, 1H), 7.19–7.41 (m, 5H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 13.9, 18.7, 21.3, 30.7, 37.9, 43.6, 55.6, 66.0, 127.4, 129.0, 129.5, 135.5, 153.2, 177.2. **MS (ESI+)** *m/z* calculated for $C_{15}H_{20}NO_3$ $[M+H]^+$: 262.32; found: 262.17.

(S)-4-Benzyl-3-((S)-2,3,3-trimethylbutanoyl)oxazolidin-2-one (**15**): Compound **15** was synthesized according to the general procedure using diisopropylamine (1383 mg, 13.66 mmol), n-BuLi (2.5 M in hexanes, 5.0 mL, 12.50 mmol), oxazolidinone **10** (2895 mg, 10.51 mmol) and MeI (5967 mg, 42.02 mmol). Recrystallization of the chromatographically pure product from cyclohexane gave the diastereomerically enriched product as a colorless solid (*dr* 20:1, 2486 mg, 8.59 mmol, 82%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 1.00 (s, 9H), 1.19 (d, *J* = 7.0 Hz, 3H), 2.76 (dd, *J* = 13.3, 9.7 Hz, 1H), 3.27 (dd, *J* = 13.3, 3.2 Hz, 1H), 3.89 (q, *J* = 7.0 Hz, 1H), 4.15 (d, *J* = 4.8 Hz, 2H), 4.62–4.74 (m, 1H), 7.19–7.42 (m, 5H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 13.2, 27.5, 33.7, 37.9, 44.8, 55.7, 65.8, 127.4, 129.0, 129.6, 135.5, 153.5, 176.8. **MS (ESI+)** *m/z* calculated for $C_{17}H_{24}NO_3$ $[M+H]^+$: 290.38; found: 290.07.

(R)-4-Benzyl-3-((R)-2-methylbutanoyl)oxazolidin-2-one (**16**): Compound **16** was synthesized according to the general procedure using diisopropylamine (1842 mg, 18.20 mmol), n-BuLi (2.5 M in hexanes, 6.7 mL, 16.80 mmol), oxazolidinone **11** (3460 mg, 14.00 mmol) and MeI (7949 mg, 56.00 mmol). The product was obtained as a colorless oil (*dr* 16:1, 2647 mg, 10.13 mmol, 72%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.94 (t, *J* = 7.4 Hz, 3H), 1.23 (d, *J* = 6.9 Hz, 3H), 1.42–1.54 (m, 1H), 1.72–1.85 (m, 1H), 2.78 (dd, *J* = 13.3, 9.6 Hz, 1H), 3.28 (dd, *J* = 13.3, 3.3 Hz, 1H), 3.60–3.69 (m, 1H), 4.15–4.24 (m, 2H), 4.65–4.72 (m, 1H), 7.20–7.37 (m, 5H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 11.8, 17.0, 26.5, 38.0, 39.3, 55.5, 66.1, 127.5, 129.0, 129.6, 135.5, 153.2, 177.3. **MS (ESI+)** *m/z* calculated for $C_{15}H_{23}N_2O_3$ $[M+H]^+$: 279.35; found: 279.24.

(R)-4-Benzyl-3-((R)-2,3-dimethylbutanoyl)oxazolidin-2-one (**17**): Compound **17** was synthesized according to the general procedure using diisopropylamine (658 mg, 6.50 mmol), n-BuLi (2.5 M in hexanes, 2.4 mL, 6.00 mmol), oxazolidinone **12** (1307 mg, 5.00 mmol) and MeI (2839 mg, 20.00 mmol). Recrystallization of the chromatographically pure product from cyclohexane gave the diastereomerically enriched product as a colorless solid (*dr* 20:1, 2486 mg, 8.59 mmol, 61%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.93 (d, *J* = 6.8 Hz, 3H), 0.95 (d, *J* = 6.7 Hz, 3H), 1.17 (d, *J* = 6.9 Hz, 3H), 1.94–2.04 (m, 1H), 2.76 (dd, *J* = 13.3, 9.7 Hz, 1H), 3.28 (dd, *J* = 13.4, 3.3 Hz, 1H), 3.59 (quint, *J* = 6.9 Hz, 1H), 4.13–4.22 (m, 2H), 4.61–4.72 (m, 1H), 7.20–7.36 (m, 5H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 13.9, 18.6, 21.3, 30.7, 37.9, 43.6, 55.5, 66.0, 127.4, 129.0, 129.5, 135.4, 153.2, 177.1. **MS (ESI+)** *m/z* calculated for $C_{16}H_{22}NO_3$ $[M+H]^+$: 276.36; found: 276.12.

(R)-4-Benzyl-3-((R)-2,3,3-trimethylbutanoyl)oxazolidin-2-one (**18**): Compound **18** was synthesized according to the general procedure using diisopropylamine (896 mg, 8.85 mmol),

n-BuLi (2.5 M in hexanes, 3.3 mL, 8.17 mmol), oxazolidinone **13** (1876 mg, 6.81 mmol) and MeI (3866 mg, 27.24 mmol). Recrystallization of the chromatographically pure product from cyclohexane gave the diastereomerically enriched product as a colorless solid (*dr* 20:1, 1379 mg, 4.77 mmol, 70%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 1.00 (s, 9H), 1.19 (d, *J* = 7.0 Hz, 3H), 2.75 (dd, *J* = 13.3, 9.7 Hz, 1H), 3.27 (dd, *J* = 13.3, 3.2 Hz, 1H), 3.88 (q, *J* = 7.0 Hz, 1H), 4.11–4.17 (m, 2H), 4.64–4.72 (m, 1H), 7.19–7.36 (m, 5H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 13.2, 27.5, 33.6, 37.9, 44.7, 55.6, 65.8, 127.4, 129.0, 129.6, 135.5, 153.5, 176.8. **MS (ESI+)** *m/z* calculated for $C_{12}H_{24}NO_3$ $[M+H]^+$: 290.38; found: 290.13.

General procedure for the synthesis of carboxylic acids **19**–**23***:* To a cooled solution of LiOH (2.0 eq) and H_2O_2 (4.0 eq) in water, a solution of acyloxazolidinone **14**–**18** (1.0 eq) in THF was added. The reaction was stirred at room temperature for 16 h. The solution was extracted with DCM (3×15 mL) and the organic phase was discarded. The aqueous phase was acidified and extracted with DCM (3×15 mL). The combined organic extracts were washed with brine and dried over $MgSO_4$. The crude reaction product was purified by silica gel column chromatography (pentane/Et_2O 5:1).

(S)-2,3-Dimethylbutanoic acid (**19**): Compound **19** was synthesized according to the general procedure using LiOH (184 mg, 4.38 mmol), H_2O_2 (30% in H_2O, 1000 mg, 8.76 mmol) and intermediate **14** (573 mg, 2.19 mmol). The product was obtained as a colorless liquid (254 mg, 2.19 mmol, 100%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.93 (d, *J* = 6.8 Hz, 3H), 0.97 (d, *J* = 6.8 Hz, 3H), 1.13 (d, *J* = 7.0 Hz, 3H), 1.88–2.02 (m, 1H), 2.22–2.33 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 13.5, 19.1, 20.8, 30.9, 46.0, 182.8. **MS (ESI-)** *m/z* calculated for $C_6H_{11}O_2$ $[M-H]^-$: 115.15; found: 115.15.

(S)-2,3,3-Trimethylbutanoic acid (**20**): Compound **20** was synthesized according to the general procedure using LiOH (203 mg, 4.84 mmol), H_2O_2 (30% in H_2O, 1097 mg, 9.68 mmol) and intermediate **15** (700 mg, 2.42 mmol). The product was obtained as a colorless liquid (313 mg, 2.40 mmol, 99%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.99 (s, 9H), 1.14 (d, *J* = 7.1 Hz, 3H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 12.7, 27.6, 32.7, 49.6, 182.5. **MS (ESI-)** *m/z* calculated for $C_7H_{13}O_2$ $[M-H]^-$: 129.18; found: 129.16.

(R)-2-Methylbutanoic acid (**21**): Compound **21** was synthesized according to the general procedure using LiOH (722 mg, 17.20 mmol), H_2O_2 (30% in H_2O, 3900 mg, 34.40 mmol) and intermediate **16** (1998 mg, 7.65 mmol). The product was obtained as a colorless liquid (760 mg, 7.44 mmol, 97%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.95 (t, *J* = 7.5 Hz, 3H), 1.18 (d, *J* = 7.0 Hz, 3H), 1.45–1.55 (m, 1H), 1.66–1.76 (m, 1H), 2.36–2.44 (m, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 11.7, 16.5, 26.7, 41.0, 183.2. **MS (ESI-)**: *m/z* calculated for $C_5H_9O_2$ $[M-H]^-$: 101.13; found: 101.11.

(R)-2,3-Dimethylbutanoic acid (**22**): Compound **22** was synthesized according to the general procedure using LiOH (258 mg, 6.14 mmol), H_2O_2 (30% in H_2O, 1393 mg, 12.28 mmol) and intermediate **17** (845 mg, 3.07 mmol). The product was obtained as a colorless liquid (332 mg, 2.86 mmol, 93%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.93 (d, *J* = 6.8 Hz, 3H), 0.96 (d, *J* = 6.8 Hz, 3H), 1.13 (d, *J* = 7.0 Hz, 3H), 1.90–2.01 (m, 1H), 2.21–2.30 (m, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 13.5, 19.1, 20.8, 30.8, 46.0, 182.8. **MS (ESI-)** *m/z* calculated for $C_6H_{11}O_2$ $[M-H]^-$: 115.15; found: 115.15.

(R)-2,3,3-Trimethylbutanoic acid (**23**): Compound **23** was synthesized according to the general procedure using LiOH (132 mg, 3.14 mmol), H_2O_2 (30% in H_2O, 713 mg, 6.28 mmol) and intermediate **18** (455 mg, 1.57 mmol). The crude reaction product was purified by silica gel column chromatography (pentane/Et_2O 5:1). The product was obtained as a colorless liquid (177 mg, 1.36 mmol, 87%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.99 (s, 9H), 1.14 (d, *J* = 7.1 Hz, 3H), 2.24–2.34 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 12.7, 27.6, 32.7, 49.5, 182.1. **MS (ESI-)** *m/z* calculated for $C_7H_{13}O_2$ $[M-H]^-$: 129.18; found: 129.22.

General procedure for the synthesis of alcohols **24**–**28***:* A solution of $LiAlH_4$ in Et_2O (2.0 eq) was added dropwise to a cooled solution of the respective carboxylic acid **19**–**23** (1.0 eq) in dry Et_2O. After stirring for 1 h the reaction was quenched by the addition of H_2O and 1 M HCl. The crude reaction mixture was filtered over Celite and dried over $MgSO_4$.

(S)-2,3-Dimethylbutan-1-ol (**24**): Compound **24** was synthesized according to the general procedure using $LiAlH_4$ (1 M in Et_2O, 4.0 mL, 4.02 mmol) and carboxylic acid **19** (233 mg, 2.01 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (161 mg, 1.58 mmol, 78%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.83 (d, *J* = 6.8Hz, 3H), 0.86 (d, *J* = 6.9 Hz, 3H), 0.91 (d, *J* = 6.9 Hz, 3H), 1.44–1.54 (m, 1H), 1.64–1.76 (m, 1H), 3.45 (dd, *J* = 10.5, 7.0 Hz, 1H), 3.59 (dd, *J* = 10.6, 5.9 Hz, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 12.6, 18.1, 20.7, 28.9, 41.5, 66.7.

(S)-2,3,3-Trimethylbutan-1-ol (**25**): Compound **25** was synthesized according to the general procedure using $LiAlH_4$ (1 M in Et_2O, 10.0 mL, 10.00 mmol) and carboxylic acid **20** (1118 mg, 8.59 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (950 mg, 8.17 mmol, 95%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.88 (s, 9H), 0.94 (d, *J* = 6.8 Hz, 3H), 1.35–1.45 (m, 1H), 3.28–3.36 (m, 1H), 3.82 (dd, *J* = 10.3, 3.7 Hz, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 12.5, 27.8, 32.2, 45.7, 65.5.

(R)-2-Methylbutan-1-ol (**26**): Compound **26** was synthesized according to the general procedure using $LiAlH_4$ (1 M in Et_2O, 7.2 mL, 7.20 mmol, 1.2 eq) and carboxylic acid **23** (613 mg, 6.00 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (493 mg, 5.60 mmol, 93%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.89–0.93 (m, 6H), 1.10–1.24 (m, 1H), 1.41–1.49 (m, 1H), 1.51–1.59 (m, 1H), 3.39–3.54 (m, 2H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 11.5, 16.3, 25.9, 37.5, 68.2.

(R)-2,3-Dimethylbutan-1-ol (**27**): Compound **27** was synthesized according to the general procedure using $LiAlH_4$ (1 M in Et_2O, 6.0 mL, 6.00 mmol) and carboxylic acid **22** (326 mg, 2.81 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (176 mg, 1.72 mmol, 61%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.83 (d, *J* = 6.8 Hz, 3H), 0.86 (d, *J* = 6.9 Hz, 3H), 0.91 (d, *J* = 6.9 Hz, 3H), 1.45–1.54 (m, 1H), 1.65–1.73 (m, 1H), 3.44 (dd, *J* = 10.6, 7.0 Hz, 2H), 3.59 (dd, *J* = 10.6, 5.9 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 12.6, 18.1, 20.7, 28.9, 41.5, 66.7.

(R)-2,3,3-Trimethylbutan-1-ol (**28**): Compound **28** was synthesized according to the general procedure using $LiAlH_4$ (1 M in Et_2O, 2.7 mL, 2.70 mmol) and carboxylic acid **23** (177 mg, 1.36 mmol). Careful evaporation of the solvent gave the product as a colorless liquid (106 mg, 0.91 mmol, 67%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.88 (s, 9H), 0.95 (d, *J* = 6.9 Hz, 3H), 1.37–1.45 (m, 1H), 3.33 (dd, *J* = 10.4, 8.8 Hz, 1H), 3.82 (dd, *J* = 10.4, 3.9 Hz, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 12.5, 27.8, 32.2, 45.8, 65.5.

General procedure for the synthesis of Intermediates **34**–**38***:* PPh_3 (3.0 eq) was dissolved in dry THF and cooled to -15 °C. DEAD (1.4 eq) was added to the solution without the temperature exceeding -10 °C. After 30 minutes the respective alcohol was added to the reaction mixture and it was stirred at -15 °C for another 90 minutes. DPPA (1.5 eq) was slowly added to the solution and the reaction was held at -15 °C for further 4 h and then allowed to warm to RT overnight. Water (3 mL) was added to the solution and the mixture was stirred for 48 h. The crude reaction mixture was diluted with EtOAc and extracted with 1 M HCl (3×10 mL). The aqueous phases were combined and the pH was adjusted to 12. The basic mixture was extracted with DCM (3×15 mL). The combined organic phases were washed with brine and dried over K_2CO_3. Boc-Leu-OH (1.0 eq), EDC (1.3 eq), HOBt (1.3 eq) and DIPEA (3.0 eq) were added to the DCM phase and the mixture was stirred at rt overnight. The solution was diluted with EtOAc and extracted with 1 M HCl (3×10 mL) and sat. $NaHCO_3$ (3×10 mL). The combined organic phases were washed with brine and dried over $MgSO_4$. The crude product was purified by silica gel column chromatography (cyclohexane/EtOAc 7:1).

tert-Butyl ((S)-1-(((S)-2,3-dimethylbutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**34**): Intermediate **34** was synthesized according to the general procedure using PPh_3 (1141 mg, 4.35 mmol), DEAD (354 mg, 2.03 mmol), alcohol **24** (148 mg, 1.45 mmol), DPPA (600 mg, 2.18 mmol), Boc-Leu-OH (335 mg, 1.45 mmol), EDC (362 mg, 1.89 mmol), HOBt (255 mg, 1.89 mmol), DIPEA (562 mg, 4.35 mmol). The product was obtained as a colorless solid (119 mg, 0.38 mmol, 26%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.80–0.85 (m, 6H), 0.88–0.95 (m, 9H), 1.43 (s, 9H), 1.46–1.79 (m, 5H), 2.99–3.29 (m, 2H), 3.92–4.15 (m, 1H), 4.84 (br, 1H), 6.12 (br, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 13.6, 17.9, 20.5, 22.2, 22.9, 24.8, 28.4, 30.0, 39.0, 41.1, 43.5, 53.2, 80.1, 155.9, 172.6. **MS (ESI+)** *m/z* calculated for $C_{17}H_{35}N_2O_3$ $[M+H]^+$: 315.48; found: 315.20.

tert-Butyl ((S)-1-(((S)-2,3,3-trimethylbutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**35**): Intermediate **35** was synthesized according to the general procedure using PPh_3 (2361 mg, 9.00 mmol), DEAD (731 mg, 4.20 mmol), alcohol **25** (345 mg, 3.00 mmol), DPPA (1238 mg, 4.50 mmol), Boc-Leu-OH (694 mg, 3.00 mmol), EDC (748 mg, 3.90 mmol), HOBt (527 mg, 3.90 mmol), DIPEA (1163 mg, 9.00 mmol). The product was obtained as a colorless solid (319 mg, 0.97 mmol, 32%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.85 (d, *J* = 6.8 Hz, 3H), 0.89 (s, 9H), 0.91–0.95 (m, 6H), 1.34–1.41 (m, 1H), 1.43 (s, 9H), 1.62–1.72 (m, 2H), 2.89–2.98 (m, 1H), 3.37–3.48 (m, 1H), 4.00–4.10 (m, 1H), 4.87 (br, 1H), 6.07–6.18 (m, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 13.1, 22.2, 23.0, 24.9, 27.5, 28.4, 32.5, 41.1, 41.9, 43.3, 53.3, 80.1, 155.9, 172.6. **MS (ESI+)** *m/z* calculated for $C_{18}H_{37}N_2O_3$ $[M+H]^+$: 329.50; found: 329.19.

tert-Butyl ((S)-4-methyl-1-(((R)-2-methylbutyl)amino)-1-oxopentan-2-yl)carbamate (**36**): Intermediate **36** was synthesized according to the general procedure using PPh_3 (2361 mg, 9.00 mmol), DEAD (731 mg, 4.20 mmol), alcohol **26** (264 mg, 3.00 mmol), DPPA (1238 mg, 4.50 mmol), Boc-Leu-OH (694 mg, 3.00 mmol), EDC (748 mg, 3.90 mmol), HOBt (527 mg, 3.90 mmol), DIPEA (1163 mg, 9.00 mmol). The product was obtained as a colorless solid (327 mg, 1.09 mmol, 36%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.85–0.96 (m, 12H), 1.08–1.20 (m, 1H), 1.32–1.40 (m, 1H), 1.43 (s, 9H), 1.49–1.61 (m, 2H), 1.61–1.73 (m, 2H), 3.02–3.23 (m, 2H), 3.99–4.11 (m, 1H), 4.84–4.94 (m, 1H), 6.19 (s, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 11.4, 17.2, 22.3, 23.0, 24.9, 27.0, 28.4, 35.0, 41.1, 45.1, 53.3, 80.1, 156.0, 172.7. **MS (ESI+)** *m/z* calculated for $C_{16}H_{33}N_2O_3$ $[M+H]^+$: 301.45; found: 301.28.

tert-Butyl ((S)-1-(((R)-2,3-dimethylbutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**37**): Intermediate **37** was synthesized according to the general procedure using PPh_3 (1251 mg, 4.77 mmol), DEAD (388 mg, 2.23 mmol), alcohol **27** (162 mg, 1.59 mmol), DPPA (658 mg, 2.39 mmol), Boc-Leu-OH (368 mg, 1.59 mmol), EDC (397 mg, 2.07 mmol), HOBt (280 mg, 2.07 mmol), DIPEA (617 mg, 4.77 mmol). The product was obtained as a colorless solid (136 mg, 0.43 mmol, 22%). **^{1}H NMR** (500 MHz, $CDCl_3$) δ = 0.81–0.85 (m, 6H), 0.90 (d, *J* = 6.8 Hz, 3H), 0.91–0.95 (m, 6H), 1.43 (s, 9H), 1.45–1.54 (m, 1H), 1.55–1.64 (m, 1H), 1.64–1.75 (m, 2H), 3.02–3.12 (m, 1H), 3.17–3.31 (m, 1H), 3.99–4.08 (m, 1H), 4.87 (br, 1H), 6.16 (br, 1H). **^{13}C NMR** (126 MHz, $CDCl_3$) δ = 13.59, 17.98, 20.57, 22.28, 22.98, 24.88, 28.41, 30.07, 38.99, 41.13, 43.53, 53.27, 80.08, 155.95, 172.63. **MS (ESI+)** *m/z* calculated for $C_{17}H_{35}N_2O_3$ $[M+H]^+$: 315.48; found: 315.05.

tert-Butyl ((S)-1-(((R)-2,3,3-trimethylbutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamate (**38**): Intermediate **38** was synthesized according to the general procedure using PPh_3 (716 mg, 2.73 mmol), DEAD (222 mg, 1.27 mmol), alcohol **28** (106 mg, 0.91 mmol), DPPA (377 mg, 1.37 mmol), Boc-Leu-OH (210 mg, 0.91 mmol), EDC (227 mg, 1.18 mmol), HOBt (159 mg, 1.18 mmol), DIPEA (353 mg, 2.73 mmol). The product was obtained as a colorless solid (80 mg, 0.24 mmol, 27%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.85 (d, *J* = 6.9 Hz, 3H), 0.88 (s, 9H),

0.90–0.96 (m, 6H), 1.32–1.40 (m, 2H), 1.43 (s, 9H), 1.58–1.75 (m, 2H), 2.88–2.97 (m, 1H), 3.38–3.49 (m, 1H), 3.99–4.14 (m, 1H), 4.86–4.97 (m, 1H), 6.17 (br, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 13.1, 22.3, 23.0, 24.9, 27.5, 28.4, 32.6, 41.1, 42.0, 43.3, 53.3, 80.2, 155.9, 172.6. **MS (ESI+)** *m/z* calculated for $C_{18}H_{37}N_2O_3$ $[M+H]^+$: 329.50; found: 329.25.

General procedure for phosphonamidate coupling and deprotection: $SOCl_2$ (4.5 eq) was dissolved in 5 mL dry DCM under argon and cooled to 0 °C. A solution of benzyl ((hydroxy(methoxy)phosphoryl)methyl)carbamate (3 eq) was added to the solution over 30 minutes. The cooling bath was removed and the reaction was stirred at rt for 3-5 h. All volatile components were removed under reduced pressure. In a second flask, peptidic intermediates **34**–**38** (1 eq) was treated with HCl (4 M in dioxane,) for 1 h. All volatile components were removed under reduced pressure. The activated phosphonic acid was dissolved in dry DCM. DIPEA and the deprotected leucylamide were added to the solution consecutively. The mixture was heated to 40 °C and stirred overnight. The solution was diluted with EtOAc and extracted with 5% citric acid (3×10 mL), 1 M HCl (1×10 mL) and 1 M NaOH (3×10 mL). The organic phase was washed with brine and dried over $MgSO_4$. The crude product was used in the next step without further purification. The protected phosphonamidate was treated with 3mL of a 0.4 M solution of LiOH in water. If needed, MeCN was added until the solution cleared. The reaction was stirred for 4-5 h at rt. Under ice cooling the pH was adjusted to 8 using 5% AcOH and the solvent was removed under reduced pressure. The residue was dissolved in a small amount of water and purified by semi-preparative HPLC.

Phosphonamidate **2***:* Ligand **2** was synthesized according to the general procedure using phosphonic acid monoester **39** (249 mg, 0.96 mmol), $SOCl_2$ (171 mg, 1.44 mmol), intermediate **34** (100 mg, 0.32 mmol), HCl (4 M in dioxane, 1 mL, 4 mmol) and DIPEA (124 mg, 0.96 mmol). The product was obtained as a colorless solid (45 mg, 0.10 mmol, 31%). **^{1}H NMR** (500 MHz, D_2O) δ = 0.84 (d, *J* = 6.8 Hz, 3H), 0.87 (d, *J* = 6.7 Hz, 3H), 0.92–0.98 (m, 9H), 1.44–1.52 (m, 1H), 1.52–1.76 (m, 4H), 3.04 (dd, *J* = 13.2, 8.0 Hz, 1H), 3.20 (dd, *J* = 13.4, 6.0 Hz, 1H), 3.26–3.40 (m, 2H), 3.67–3.75 (m, 1H), 5.20 (s, 2H), 7.45–7.56 (m, 5H). **^{13}C NMR** (126 MHz, D_2O) δ = 12.7, 17.1, 19.8, 21.5, 22.3, 24.3, 29.5, 38.3, 39.9 (d, *J* = 136.0), 43.5, 43.7, 43.7, 54.3, 67.3, 127.9, 128.5, 128.9, 136.5, 158.3, 177.8. **^{31}P NMR** (202 MHz, D_2O) δ = 18.0. **HRMS (ESI+)** calculated for $C_{21}H_{35}N_3O_5P$ $[M+H]^+$: 440.2320; found: 440.2323.

Phosphonamidate **3***:* Ligand **3** was synthesized according to the general procedure using phosphonic acid monoester **39** (207 mg, 0.80 mmol), $SOCl_2$ (190 mg, 1.60 mmol), intermediate **35** (131 mg, 0.40 mmol), HCl (4 M in dioxane, 1 mL, 4 mmol) and DIPEA (155 mg, 1.20 mmol). The product was obtained as a colorless solid (68 mg, 0.15 mmol, 37%). **^{1}H NMR** (500 MHz, D_2O) δ = 0.75 (d, *J* = 6.9 Hz, 3H), 0.81 (s, 9H), 0.82–0.86 (m, 6H), 1.30–1.41 (m, 2H), 1.42–1.52 (m, 1H), 1.53–1.65 (m, 1H), 2.84 (dd, *J* = 12.5, 11.0 Hz, 1H), 3.22 (dd, *J* = 11.8, 5.7 Hz, 2H), 3.27 (dd, *J* = 13.1, 3.4 Hz, 1H), 3.55–3.64 (m, 1H), 5.09 (s, 2H), 7.33–7.44 (m, 5H). **^{13}C**

NMR (126 MHz, D_2O) δ = 12.4, 21.5, 22.3, 24.3, 26.7, 31.8, 40.0 (d, *J* = 136.3), 42.0, 42.6, 43.7, 43.7, 54.3, 67.2, 127.8, 128.5, 128.9, 136.5, 158.3, 177.8. **^{31}P NMR** (202 MHz, D_2O) δ = 18.0. **HRMS (ESI+)** calculated for $C_{22}H_{37}N_3O_5P$: 454.2476 $[M+H]^+$; found: 454.2478.

Phosphonamidate **4***:* Ligand **4** was synthesized according to the general procedure using phosphonic acid monoester **39** (117 mg, 0.45 mmol), $SOCl_2$ (81 mg, 0.68 mmol), intermediate **36** (45 mg, 0.15 mmol), HCl (4 M in dioxane, 1 mL, 4 mmol) and DIPEA (58 mg, 0.45 mmol). The product was obtained as a colorless solid (25 mg, 0.06 mmol, 37%). **^{1}H NMR** (500 MHz, MeOD) δ = 0.85–0.98 (m, 12H), 1.09–1.19 (m, 1H), 1.35–1.48 (m, 2H), 1.52–1.63 (m, 2H), 1.72–1.82 (m, 1H), 2.93–3.04 (m, 1H), 3.09 (dd, *J* = 13.2, 6.3 Hz, 1H), 3.18–3.29 (m, 2H), 3.67–3.76 (m, 1H), 5.04–5.14 (m, 2H), 7.25–7.41 (m, 5H). **^{13}C NMR** (126 MHz, MeOD) δ = 11.6, 17.5, 22.5, 23.5, 25.7, 28.1, 36.1, 41.2, 42.3, 45.3 (d, *J* = 5.3), 46.2, 55.5, 67.6, 128.9, 129.0, 129.4, 138.3, 158.8, 178.2. **^{31}P NMR** (202 MHz, MeOD) δ = 16.4. **HRMS (ESI+)** calculated for $C_{20}H_{33}N_3O_5P$: 426.2163 $[M+H]^+$; found: 426.2166.

Phosphonamidate **5***:* Ligand **5** was synthesized according to the general procedure using phosphonic acid monoester **39** (249 mg, 0.96 mmol), $SOCl_2$ (171 mg, 1.44 mmol), intermediate **37** (100 mg, 0.32 mmol), HCl (4 M in dioxane, 1 mL, 4 mmol) and DIPEA (124 mg, 0.96 mmol). The product was obtained as a colorless solid (37 mg, 0.08 mmol, 26%). **^{1}H NMR** (500 MHz, D_2O) δ = 0.84 (d, *J* = 6.8 Hz, 3H), 0.87 (d, *J* = 6.7 Hz, 3H), 0.92–0.98 (m, 9H), 1.44–1.77 (m, 5H), 3.04 (dd, *J* = 13.3, 8.0 Hz, 1H), 3.20 (dd, *J* = 13.3, 6.0 Hz, 1H), 3.27–3.41 (m, 2H), 3.67–3.74 (m, 1H), 5.20 (s, 2H), 7.45–7.55 (m, 5H). **^{13}C NMR** (126 MHz, D_2O) δ = 12.7, 17.1, 19.8, 21.5, 22.3, 24.3, 29.5, 38.3, 39.9 (d, *J* = 135.9), 43.5, 43.7, 43.7, 54.3, 67.3, 127.9, 128.5, 128.9, 136.5, 158.3, 177.8. **^{31}P NMR** (202 MHz, D_2O) δ = 18.0. **HRMS (ESI+)** calculated for $C_{21}H_{35}N_3O_5P$ $[M+H]^+$: 440.2322; found: 440.2323.

Phosphonamidate **6***:* Ligand **6** was synthesized according to the general procedure using phosphonic acid monoester **39** (156 mg, 0.60 mmol), $SOCl_2$ (107 mg, 0.90 mmol), intermediate **38** (65 mg, 0.20 mmol), HCl (4 M in dioxane, 1 mL, 4 mmol) and DIPEA (78 mg, 0.60 mmol). The product was obtained as a colorless solid (29 mg, 0.06 mmol, 31%). **^{1}H NMR** (500 MHz, D_2O) δ = 0.75 (d, *J* = 6.9 Hz, 3H), 0.82 (s, 9H), 0.83–0.87 (m, 6H), 1.32–1.41 (m, 2H), 1.43–1.50 (m, 1H), 1.55–1.64 (m, 1H), 2.84 (dd, *J* = 12.9, 10.7 Hz, 1H), 3.18–3.31 (m, 3H), 3.56–3.63 (m, 1H), 5.09 (s, *J* = 16.7 Hz, 2H), 7.33–7.44 (m, 5H). **^{13}C NMR** (101 MHz, D_2O) δ = 12.2, 21.3, 22.1, 24.1, 26.5, 31.6, 39.8 (d, *J* = 135.7), 41.8, 42.5, 43.6, 54.2, 67.1, 127.7, 128.4, 128.8, 136.2, 152.8, 172.5. **^{31}P NMR** (162 MHz, D_2O) δ = 17.9. **HRMS (ESI+)** calculated for $C_{22}H_{37}N_3O_5P$: 454.2476 $[M+H]^+$; found: 454.2481.

3.6.3 Crystal Preparation and Soaking

TLN crystals were prepared similarly to the procedure as previously described [156]. Lyophilized TLN powder was commercially obtained from Calbiochem (EMD Biosciences). For crystal preparation, 1 mL of demineralized water was pipetted into the reservoir wells of a 24 well sitting drop crystallization plate (Cryschem, Hampton research). An 8 mM protein suspension was prepared by adding TLN powder to 50 µL of pure DMSO. To the resulting suspension, 50 µL of an aqueous solution containing 3.8 M CsCl and 100 mM Tris-HCl, pH 7.5 were added. The TLN powder was completely dissolved by mixing with an Eppendorf pipette, resulting in a clear solution of yellowish color. After centrifugation for 3 min (RCF = 16000 g), 1 µL of clear solution was transferred into each of the protein wells of the crystal plate. Subsequently, the crystal plate was sealed and TLN crystals stopped growing after five days at 18 °C. For ligand soaking, crystals with an obelisk shape were transferred into a soaking solution composed of 100 mM Tris-HCl, pH 7.5, 2 mM $CaCl_2$, 5% DMSO and 1 mM of the respective ligand followed by incubation for 24 h. Afterwards, crystals were transferred into a cryo buffer composed of 10 mM Tris/HCl, pH 7.5, 10 mM $Ca(CH_3COO)_2$, 5% DMSO, 20% glycerol and 1 mM of the respective ligand for 5 s and subsequently flash-frozen in liquid nitrogen. For the glycerol-free crystal structures (Chapter 3.10.2, Supporting Information), either 30% PEG400 or 20% MPD was used as cryoprotectant instead of glycerol.

3.6.4 Data Collection and Refinement

Data collection of datasets TLN-$\mathbf{2}_{GOL}$, TLN-$\mathbf{3}_{GOL}$, TLN-$\mathbf{4}_{GOL}$, TLN-$\mathbf{5}_{GOL}$,TLN-$\mathbf{6}_{GOL}$, TLN-$\mathbf{3}_{PEG400}$, and TLN-$\mathbf{6}_{MPD}$ was performed with synchrotron radiation at the BESSY II electron storage ring (Berlin-Adlershof, Germany) operated by the Joint Berlin MX-Laboratory [157] at beamline 14.1 with a wavelength of 0.91841 Å at cryogenic temperature of 100 K on an Dectris Pilatus 6M pixel detector. Dataset TLN-$\mathbf{5}_{MPD}$ was collected at Elettra (Trieste, Italy) at beamline XRD1 with a wavelength of 1.0000 Å at cryogenic temperature of 100 K on a Dectris Pilatus 2M pixel detector. Data collection and refinement statistics for the glycerol-containing crystal structures are listed in Table 3.1 and for the glycerol-free crystal structures in Table S3.1 (Supporting Information). Data indexing, integration and scaling was performed with *XDS* [158]. Molecular replacement was performed with *Phaser* [159] from the *CCP*4 software suite version 6.3.0 [137]. A TLN search model from the PDB entry 8TLN was used [160], with flexible side-chains, additives, water molecules and ions removed. Alternating cycles of model building and refinement were performed with *Coot* [161] and *phenix.refine* version 1.10.1-2155 [162], respectively, until the models optimally explained the electron densities and their *R* values reached convergence. For cross-validation, a randomly chosen 5% of the reflections, which were not applied for model refinement, were used for the calculation of R_{free}. As an initial

refinement step, Cartesian simulated annealing was performed (start temperature 5000 K, final temperature 300 K). All macromolecule amino acids were refined with riding hydrogen atoms which were added to the refinement model coordinates with *phenix.refine*. In all structures, *B* factors of all atoms except for hydrogen atoms were refined anisotropically, since this resulted in a significant decrease of the R_{free} value compared to an isotropically refinement of the *B* factors. Ligand molecule building and minimization was performed with *MOE* version 2014.09 [151], and restraints were prepared with *phenix.elbow* [163]. F_o–F_c omit electron densities were created by deletion of parts of the refinement model followed by refinement with *phenix.refine*. The graphical representations of the three dimensional structures were prepared with *PyMOL* [164].

3.6.5 Isothermal Titration Calorimetry Measurements

ITC titrations were performed based on established protocols introducing some modifications [15,93,94,96]. For all measurements, the same Microcal ITC_{200} device (GE Healthcare) was used. Lyophilized powder of native TLN expressed by *Bacillus thermoproteolyticus* was bought from Calbiochem (EMD Biosciences). The protein purity was 60%, additionally containing $Ca(CH_3COO)_2$ and $Na(CH_3COO)$. The powder was weighed with a MX5 balance from Mettler Toledo (readability ± 1 µg) and directly dissolved in measurement buffer without further processing, as the additional salts did not show any interfering heat effects during the measurement [15]. All measurements were performed in a buffer composed of 200 mM NaSCN, 2 mM $CaCl_2$ and 20 mM HEPES, pH 7.5, which was filtered through a 0.22 µm filter and degassed prior to use. A TLN concentration of 30 µM was used in the titration cell, resulting in titration curves described by optimal *c* values in the range of 73 (ligand **4**) to 201 (ligand **3**). The relatively narrow affinity range of the studied ligands made it possible to keep the TLN concentration constant during all measurements, which is important to guarantee a similar activity coefficient of the protein in solution [96]. The ligand solution was prepared by directly weighing the highly pure, freeze dried ligand powder and dissolving it in measurement buffer (without the addition of DMSO). For each measurement, new freeze-dried protein and ligand powder was freshly dissolved in measurement buffer. For highest measurement precision, a 10 injections scheme with an injection volume of 1.3–1.4 µL and a ligand concentration in the syringe of 1 mM was applied [165]. This measurement protocol resulted in injections with strong heat signals exhibiting lower standard deviations compared to a conventional 25 injection scheme. At the end of the titration, a titrant to titrand ratio of at least 2.2 was achieved. After filling of the syringe with ligand solution and prior to injection of the syringe into the measurement cell, a manual 1.0 µL injection was performed outside of the measurement cell in order to adjust the syringe drive mechanism from the a “up” to a “down” movement in order to increase the precision of the first injection [166]. Furthermore, the

measurement scheme contained an initial injection of 0.3 µL, which was excluded from data analysis. The measurement settings were adjusted to an initial spacing of 150–200 s, a spacing of 150–160 s between the injections, an injection speed of 2.6–2.8 s (depending on the injection volume), a syringe speed of 1000 rpm, a reference power of 5 µcal s^{-1} and a measurement temperature of 298.15 K. Peak extraction and integration was automatically performed with *NITPIC* version 1.1.2 [167]. The fitting of a 1:1 binding model curve and binding parameter extraction was performed with *SEDPHAT* version 12.1b [168] and plots of the raw data and binding isotherms were prepared with *GUSSI* [169] (Figure S3.5, Supporting Information). For the study of the influence of glycerol on the thermodynamic binding profiles of **2**, **3**, **5** and **6**, buffers composed of 200 mM NaSCN, 2 mM $CaCl_2$, 20 mM HEPES, pH 7.5 and a glycerol concentration of 1%, 2.5%, 5% or 10% were used. For the titrations of **3** and **6** with the addition of DMSO, buffers composed of 200 mM NaSCN, 2 mM $CaCl_2$, 20 mM HEPES, pH 7.5 and a DMSO concentration of 0.065 M or 0.130 M DMSO were used. All measurements were performed in triplicate, from which the mean and the standard deviation were calculated.

3.6.6 **Surface Plasmon Resonance Measurements**

SPR measurements were performed on a Biacore T200 system (GE Healthcare, Sweden) with analysis and sample compartment temperature set to 25 °C. The binding kinetic assay was developed as a capture assay of biotinylated TLN utilizing the Biotin CAPture Kit (GE Healthcare) combined with single cycle kinetics of the inhibitors. For the biotinylation process, lyophilized powder of native TLN (Calbiochem, see above) was dissolved to approximately 0.5 µg mL^{-1} in 0.1 M sodium borate buffer pH 8.5 supplemented with 2 mM $CaCl_2$ and preincubated with excess molar concentration of a high affinity TLN-inhibitor to protect the binding site from biotinylation. The biotinylation reaction was set up using the EZ-Link Sulfo-NHS-LC-Biotin reagent (Pierce) with a 0.6 biotin/protein molar ratio. Following incubation for 1 h at room temperature, the labelled protein was purified from non-reacted biotin reagents by desalting on a PD-10 column according to the manufacturer's instructions (GE Healthcare) to 50 mM Tris pH 8.0, 100 mM NaSCN, 2 mM $CaCl_2$. Concentration of purified TLN-biotin was determined using the mean value of triplicate absorbance readings at 280 nm on a ND1000 spectrometer (Nanodrop). Aliqous of TLN-biotin were kept at 4 °C for use within the next two days, alternatively flash frozen in liquid nitrogen for storage at -70 °C. Before use, TLN-biotin was defrosted on ice if frozen and centrifuged in a bench top centrifuge at high speed for 5 min at 4 °C.

Sensor Chip CAP was docked in the Biacore and prepared according to the manufacturer's instructions. For each day of SPR measurements, new dilutions of inhibitors from 10 mM stock solutions in 100% DMSO (stored at -20 °C) and TLN-biotin, were made using freshly prepared buffer. 50 mM Tris pH 8.0, 100 mM NaSCN, 2 mM $CaCl_2$, 2% DMSO was used as sample and

running buffer. Runs included three start-up cycles and each analysis cycle contained 4 steps: (1) injection of Biotin CAPture reagent (streptavidin modified with a deoxyriboologonucletide that hybridizes with the complimentary oligonucleotide present on the Sensor Chip CAP) in both reference and active flow cells for 300 s and at 2 µl min^{-1} resulting in response levels of around 3000 RU, (2) injection of 150–200 µg mL^{-1} TLN-biotin at 10 µL min^{-1} for 90–180 s in active flow cell only resulting in capture levels of 800–1200 RU (a capture stabilization time of 300 s was applied for some runs), (3) injection of inhibitor in increasing concentration over reference and active flow cell using a single cycle kinetics procedure with five 120 s injections at 30 µL min^{-1} and a 240 s dissociation time after the final injection, and (4) injection of standard (6 M guanidine-HCl, 0.25 M NaOH) and additional (30% acetonitrile in 0.25 M NaOH) regeneration solutions to remove the Biotin Cap Reagent, TLN-biotin and any bound inhibitor from both flow cells. Inhibitor concentrations in the range of 156–2500 nM were used for **1**, 25–2000 nM for **2** and **5** and 156–10000 nM for **3** and **6**. Blank cycles defined as analysis cycles with buffer only in step (3), were performed first, last and between every inhibitor concentration series. Data were double referenced by first subtraction of reference flow cell and then subtraction of blank cycles. Fitting of data was performed using Biacore T200 evaluation software 2.0, applying a 1:1 binding model compensating for linear drift.

3.6.7 Calculation of Buried Solvent Accessible Surface Areas

The total and buried SASAs of **1**–**6** (Figure 3.10) were calculated with the PISA server from the European Bioinformatics Institute [170]. The buried solvent accessible surface area is defined as the SASA of the ligand which becomes inaccessible to water molecules (radius 1.4 Å) through binding of the ligand to the protein cavity. From the disordered ligand's carbamate group, only the orientation with the carbonyl oxygen directed to the right (relative to the view of Figure 3.4) was considered for the calculation of the surface. Furthermore, the zinc ion, crystallographic additives (glycerol, DMSO) and the crystallographic symmetry mate were considered for the calculations, whereas hydrogen atoms were excluded from the calculation.

ASSOCIATED CONTENT

3.7 Accession Codes

All described crystal structures were deposited in the PDB [171] and are available upon article publication under the accession codes 5JT9 (TLN-$\mathbf{2}_{GOL}$), 5JS3 (TLN-$\mathbf{3}_{GOL}$), 5JXN (TLN-$\mathbf{4}_{GOL}$), 5JVI (TLN-$\mathbf{5}_{GOL}$), 5JSS (TLN-$\mathbf{6}_{GOL}$), 5L8P (TLN-$\mathbf{3}_{PEG400}$), 5L41 (TLN-$\mathbf{5}_{MPD}$), and 5L3U (TLN-$\mathbf{6}_{MPD}$).

AUTHOR INFORMATION

Corresponding Author

*Phone: +49 6421 28 21313. E-mail: klebe@staff.uni-marburg.de

3.8 Author Contributions

‖S.G.K. and J.C. contributed equally to this work. The study was designed by S.G.K., J.C., M.B. and G.K. M.B. performed the MD simulations. J.C. synthesized the ligands. S.G.K. and A.H. determined the crystal structures. S.G.K. collected the thermodynamic data. V.F. performed the kinetic experiments. Kinetic data was analyzed by V.F. and R.K. Data interpretation was performed by S.G.K, J.C. and G.K. This manuscript was written by S.G.K, J.C. and G.K. with contributions from M.B. and V.F.

3.9 Acknowledgement

The authors want to thank the MX-teams at BESSY II (Helmholtz-Zentrum Berlin, Germany) and at Elettra Sincrotrone (Trieste, Italy) for their advice during data collection. The authors acknowledge the receipt of a travel grant from the Helmholtz-Zentrum Berlin (Germany). This work was funded by the European Research Council (ERC) of the European Union (grant 268145-DrugProfilBind). The authors thank Dr. Alexander Metz (Univ. of Marburg) for proof-reading of the manuscript.

3.10 Supporting Information

3.10.1 F_o–F_c omit electron densities of TLN-bound ligands 1–6

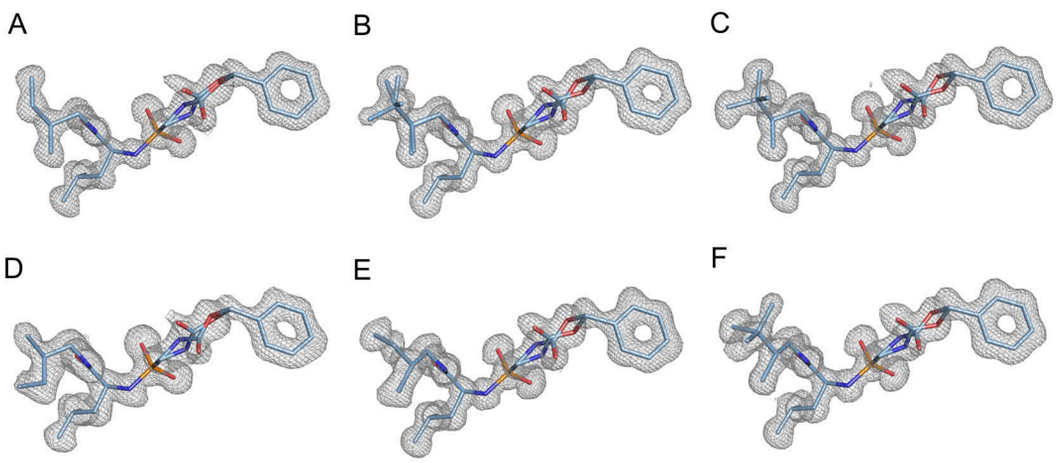


Figure S3.1 (related to Figure 3.4). Crystal structures of TLN bound ligands **(A) 1**, **(B) 2**, **(C) 3**, **(D) 4**, **(E) 5**, and **(F) 6**. Ligands are shown as stick models with carbon atoms in *blue* and color-coded heteroatoms. Their F_o–F_c omit electron densities are displayed as *gray* meshes at a contour level of 3σ. In all six crystal structures, the carbamate group of the bound ligands adopts two conformations. Moreover, the P_2' group of **2** adopts two conformations and one terminal methyl group of the P_2' group of **5** is not detectable in the electron density. The crystal structure of **1** has been published previously [15].

3.10.2 Glycerol-free crystal structures TLN-3_{PEG400}, TLN-5_{MPD} and TLN-6_{MPD}

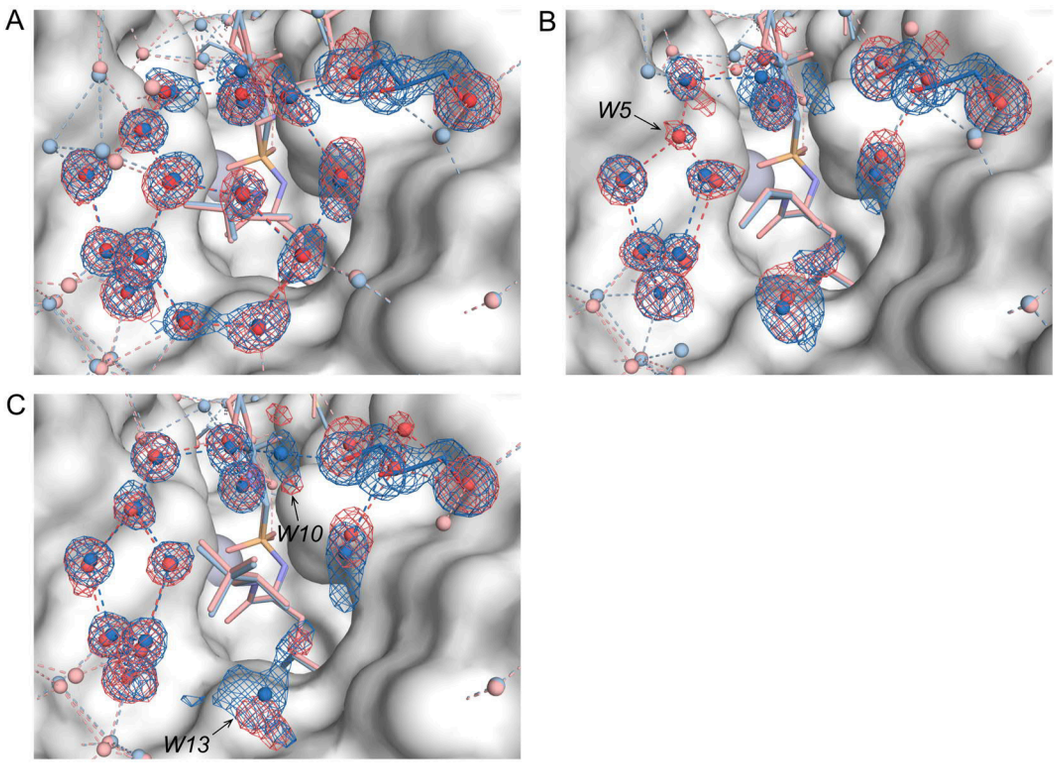


Figure S3.2. Comparison of the positions of the water molecules in the glycerol-containing and glycerol-free crystal structures of **(A)** TLN-**3**, **(B)** TLN-**5** and **(C)** TLN-**6**. Structural elements of the glycerol-containing crystal structures (*blue*) are superimposed on structural elements of the glycerol-free crystal structures (*red*). Ligand and glycerol molecules are displayed as stick models, water molecules as spheres. Distances ≤ 3.4 Å between water molecules are displayed as dotted lines indicating H-bonds. The F_o–F_c omit electron densities are shown as meshes at a contour level of 3σ in colors corresponding to the structures. For clarity reason, structural elements beyond the first solvation layer around all ligand P_2' groups and the glycerol molecules are displayed in pale colors and their electron densities are omitted. The solvent excluded surface of the glycerol-containing TLN crystal structure is displayed in *white*. The binding modes of the ligands in the glycerol-containing and glycerol-free crystal are identical and only minor differences in the adjacent water structures are observed. In TLN-$\mathbf{5}_{MPD}$ (panel B, *red*), water molecule *W5* is observed in the electron density (highly mobile, non-normalized *B* factor of 49 Å^2), whereas the electron density of TLN-$\mathbf{5}_{GOL}$ (panel B, *blue*) is too weak for the placement of a water molecule at this position in the refinement model. Furthermore, in TLN-$\mathbf{6}_{MPD}$ (panel C, *red*), the electron density is too weak for the placement of water molecules *W10* and *W13*, whereas in TLN-$\mathbf{6}_{GOL}$ (panel C, *blue*) these two water molecules are sufficiently stabilized for placement in the refinement model (highly mobile, see Figure 3.5 of the main text). F_o–F_c omit electron densities of the TLN-bound ligands of the glycerol-free crystal structures are displayed in Figure S3.3.

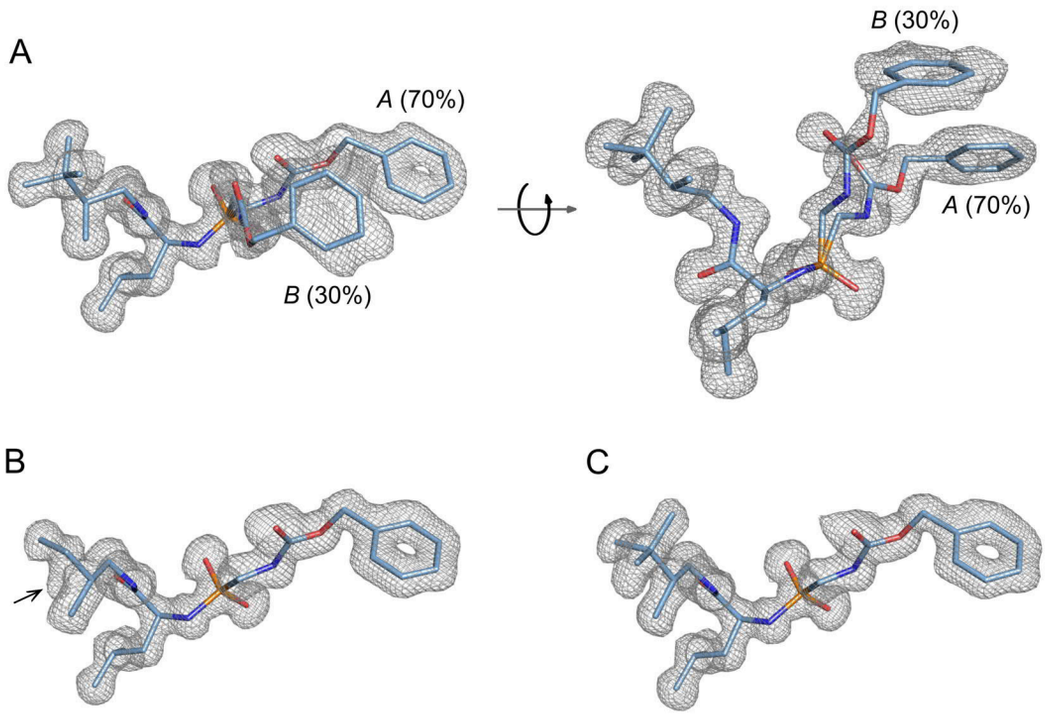


Figure S3.3. Crystal structures of the TLN-bound inhibitors **(A)** TLN-$\mathbf{3}_{PEG400}$, **(B)** TLN-$\mathbf{5}_{MPD}$ and **(C)** TLN-$\mathbf{6}_{MPD}$. Ligands are shown as stick models with carbon atoms in *blue* and color-coded heteroatoms. The F_o–F_c omit electron densities are displayed as *gray* meshes at contour levels of 2.5σ for **3** and 3.0σ for **5** and **6**. The carboxybenzyl portion of **3** shows a second conformation *B* (30% occupancy), occupying the space where a DMSO molecule is found in the other crystal structures (Figure S3.4). After refinement of the model of TLN-**3** against the diffraction data, a positive, featureless F_o–F_c electron density blob remains in position of the $2F_o$–F_c electron density of the phenyl ring of the carboxybenzyl group in conformation *B*. Since the electron density reflects the average of all conformations that a structural element adopts in the protein crystal, the unexplained F_o–F_c electron density at the position of the phenyl ring most likely originates from a DMSO molecule binding to this site in case **3** adopts conformation *A*. The electron density of ligand **5** indicates the missing terminal P_2' methyl group by a weak F_o–F_c electron density (black arrow), which, however, is not sufficient for placement of this methyl group in the refinement model.

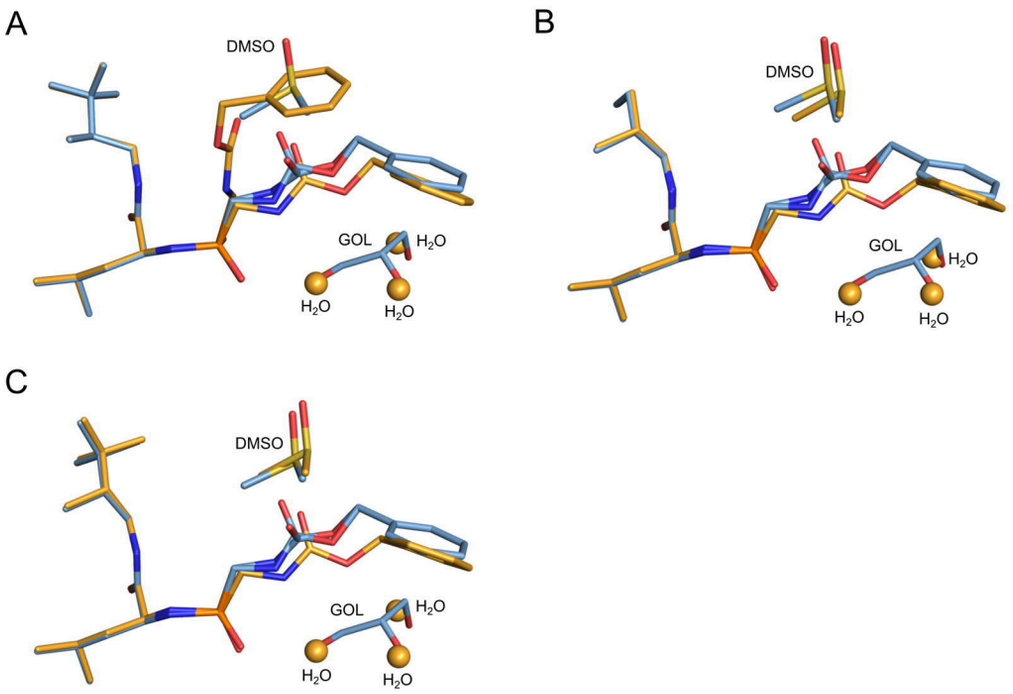


Figure S3.4. Superimposition of crystal structures of **(A)** TLN-**3**$_{GOL}$ (*blue*) and TLN-**3**$_{PEG400}$ (*orange*), **(B)** TLN-**5**$_{GOL}$ (*blue*) and TLN-**5**$_{MPD}$ (*orange*), and **(C)** TLN-**6**$_{GOL}$ (*blue*) and TLN-**6**$_{MPD}$ (*orange*). Heteroatoms are color-coded. The three spheres superimposed on the glycerol (GOL) hydroxyl groups represent water molecules bound to the TLN S_1 pocket in the superimposed glycerol-free crystal structure. As a result of the reduced steric requirement of the three water molecules compared to the glycerol molecule, the carboxybenzyl portion of the ligands in the glycerol-free crystal structures is buried more deeply in the S_1 pocket of TLN and adopts only a single, fully occupied conformation. Nevertheless, the binding mode of the leucine and P_2' portions of the ligands in the glycerol-containing and glycerol-free crystal structures are completely identical and consequently exert similar influences on the water network establishment in the TLN S_2' pocket.

Table S3.1. Data collection and refinement statistics for the glycerol-free crystal structures TLN-**2**$_{PEG400}$, TLN-**5**$_{MPD}$ and TLN-**6**$_{MPD}$.[a]

	Complex (PDB code)		
	TLN-**3**$_{PEG400}$ (5L8P)	TLN-**5**$_{MPD}$ (5L41)	TLN-**6**$_{MPD}$ (5L3U)
(A) Data collection and processing			
Space group	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters			
a, *b*, *c* (Å)	92.6, 92.6, 130.4	92.6, 92.6, 130.8	92.8, 92.8, 131.1
Matthews coefficient (Å^3/Da)[b]	2.3	2.3	2.4
Solvent content (%)[b]	47	48	48
(B) Diffraction data			
Resolution range (Å)	50.00–1.29 (1.37–1.29)	50.00–1.25 (1.33–1.25)	50.00–1.23 (1.30–1.23)
Unique reflections	83387 (13109)	91628 (14590)	96812 (15297)
$R(I)_{sym}$ (%)	6.2 (49.4)	8.0 (49.4)	5.5 (47.9)
Wilson *B* factor (Å^2)	9.6	9.3	9.4
Completeness (%)	99.7 (98.5)	100.0 (99.8)	99.8 (99.0)
Redundancy	12.7 (12.5)	12.8 (12.4)	9.6 (9.7)
$<I/\sigma(I)>$	27.6 (5.5)	19.8 (4.9)	26.8 (4.7)
(C) Refinement			
Resolution range (Å)	40.10–1.29	43.65–1.25	38.42–1.23
Reflections used in refinement (work/free)	79216/4170	87046/4582	91963/4841
Final *R* value for all reflections (work/free) (%)	10.7/13.3	10.7/13.0	10.4/12.4
Protein residues	316	316	316
Calcium/zinc ions	4/1	4/1	4/1
Inhibitor atoms	31	29	31
Water molecules	423	433	437
RMSD from ideality			
Bond lengths (Å)	0.009	0.011	0.009
Bond angles (°)	1.2	1.1	1.0
Ramachandran plot[c]			
Residues in most favored regions (%)	88.1	89.6	88.5
Residues in additionally allowed regions (%)	10.7	9.3	10.4
Residues in generously allowed regions (%)	0.7	0.7	0.7
Residues in disallowed regions (%)[d]	0.4	0.4	0.4
Mean B factor (Å^2)[e]			
Protein non-hydrogen atoms	10.7	10.6	10.7
Protein Cα atoms	9.5	9.4	9.4
Inhibitor	9.9	11.0	11.5
Water molecules	26.5	26.3	26.5

[a] Numbers in parentheses represent the values of the highest resolution shells. [b] Matthews coefficient and solvent content were calculated with the program *Matthews_coef* from the *CCP*4 suite [137]. [c] Ramachandran plots were calculated with *PROCHECK* [138]. [d] The Ramachandran outlier is Thr26 as described in literature [139]. [e] Mean *B* factors were calculated with *MOLEMAN* [140].

3.10.3 **Crystallographically determined and normalized *B* factors**

Table S3.2 (related to Figure 3.5). Crystallographically determined (non-normalized) *B* factors of *W1–W15* of TLN-**1** to TLN-**6**.

	B factor (Å^2)					
Water molecule ID	TLN-**1**	TLN-**2**	TLN-**3**	TLN-**4**	TLN-**5**	TLN-**6**
W1	11	10	10	17	10	10
W2	14	13	13	19	12	10
W3	41	25[a]	45	44	42	25
W4	-	-	32	-	49	20
W5	27	27	25	-	-	35
W6	13	13	13	23	19	17
W7	14	14	14	-	35	22
W8	31	29	19	-	47	36
W9	34	36	26	-	43	42
W10	32	34	47	49	-	49
W11	21	21	22	32	29	27
W12	38	48	41	-	-	-
W13	49	46	29	-	39	40
W14	-	-	50	-	-	-
W15	-	-	47	-	-	-

[a] both conformations *A* and *B* of *W3* in TLN-**2** have similar *B* factors.

Table S3.3 (related to Figure 3.5). Normalized *B* factors of *W1–W15* of TLN-**1** to TLN-**6**.

	normalized *B* factor (Å^2)					
Water molecule ID	TLN-**1**	TLN-**2**	TLN-**3**	TLN-**4**	TLN-**5**	TLN-**6**
W1	13	14	13	17	13	14
W2	17	17	17	19	16	14
W3	49	34[a]	59	44	57	35
W4	-	-	42	-	68	28
W5	32	37	34	-	-	48
W6	15	18	17	23	27	23
W7	17	20	18	-	48	31
W8	37	39	26	-	65	50
W9	41	49	35	-	59	58
W10	39	46	62	49	-	67
W11	26	28	30	32	40	38
W12	45	65	55	-	-	-
W13	59	62	38	-	53	55
W14	-	-	66	-	-	-
W15	-	-	63	-	-	-

[a] both conformations *A* and *B* of *W3* in TLN-**2** have similar *B* factors.

3.10.4 ITC measurements

Buffer ionization reaction during TLN–ligand complex formation

All ITC measurements were performed in HEPES buffer. Measurements conducted earlier with ligands exhibiting a similar parent scaffold (Figure 3.1A of the main text) revealed the uptake of one proton by Glu143 during the protein–ligand complex formation [93,96]. The proton is transferred from a buffer molecule, resulting in the ionization of the buffer molecule and a heat of ionization $\Delta H°_{ion}$ in addition to the heat signal $\Delta H°_{bind}$ from the actual binding event. Thus, the heat signal observed by the ITC experiment ($\Delta H°_{obs}$) is the sum of both, $\Delta H°_{ion}$ and $\Delta H°_{bind}$. Since the magnitude of the heat signal $\Delta H°_{ion}$ depends on the applied buffer molecule [123], it would be unreasonable to discuss the magnitude of $\Delta H°_{obs}$ (and thus also the calculated value of $-T\Delta S°$) on an *absolute* scale without prior correction of $\Delta H°_{ion}$. A correction of $\Delta H°_{ion}$ is possible by measuring the binding reaction in different buffers showing different heats of ionization [5,96]. However, in a congeneric series, the discussion of *relative* differences of $\Delta H°_{obs}$, all with an identical contribution of $\Delta H°_{ion}$, is possible and probably even more accurate than the comparison of calculated buffer corrected values due to error propagation. This scenario is given in the current ligand series under investigation [93], as the congeneric ligands exhibit only small changes in the aliphatic portion sticking into the solvent, and consequently exert similar heats of ionization $\Delta H°_{ion}$. As only relative differences between the thermodynamic parameters of **1**–**6** are discussed and not their absolute values, ITC measurements were only performed in one buffer and the heat of ionization was not corrected.

Remeasurement of ligand 1 to guarantee high comparability of the thermodynamic parameters

Ligand **1** was already thermodynamically characterized in the study conducted earlier [15]. However, the protein batch was completely used up for the ITC measurements and therefore it was necessary to perform the ITC measurements of the current study with TLN from a different protein batch. Furthermore, the ITC measurement scheme was optimized in order to improve measurement precision. Moreover, instead of applying *Origin7* from OriginLab for titration curve analysis, we decided to work with the programs *NITPIC* [167] and *SEDPHAT* [168] for a more automated and therefore potentially less user-biased integration and curve fitting procedure. Since ITC measurements can be very sensitive to changing measurement conditions and it is highly recommended to keep measurement conditions similar to achieve high comparability [96], we decided to measure ligand **1** again together with ligands **2**–**6**.

Table S3.4 (related to Figure 3.8). Thermodynamic parameters measured for ligands **1**–**6** by ITC. All measurements were performed in triplicate (a–c) out of which the mean values and the standard deviations were calculated. All parameters are given as rounded numbers. Measurement curves are displayed in Figure S3.5 in order to proof the high data quality of the measurements [96,172].

Measurement	*n*	K_d (μM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
TLN-**1**a	1.015	0.368	-36.7	-23.0	-13.7
TLN-**1**b	0.988	0.338	-36.9	-22.8	-14.1
TLN-**1**c	0.952	0.347	-36.9	-23.1	-13.7
	0.985±0.032	**0.351±0.015**	**-36.8±0.1**	**-23.0±0.2**	**-13.8±0.2**
TLN-**2**a	0.990	0.232	-37.9	-21.6	-16.2
TLN-**2**b	0.996	0.242	-37.8	-21.1	-16.6
TLN-**2**c	1.009	0.227	-37.9	-21.1	-16.8
	0.998±0.010	**0.233±0.008**	**-37.9±0.1**	**-21.3±0.3**	**-16.6±0.3**
TLN-**3**a	0.957	0.149	-39.0	-21.1	-17.9
TLN-**3**b	1.029	0.218	-38.0	-20.8	-17.2
TLN-**3**c	0.940	0.189	-38.4	-21.8	-16.6
	0.975±0.047	**0.185±0.035**	**-38.5±0.5**	**-21.2±0.5**	**-17.2±0.6**
TLN-**4**a	0.987	0.409	-36.5	-18.4	-18.0
TLN-**4**b	1.058	0.376	-36.7	-18.4	-18.3
TLN-**4**c	1.021	0.378	-36.7	-18.6	-18.0
	1.022±0.036	**0.388±0.019**	**-36.6±0.1**	**-18.5±0.1**	**-18.1±0.1**
TLN-**5**a	0.947	0.285	-37.4	-16.6	-20.8
TLN-**5**b	0.982	0.291	-37.3	-17.0	-20.3
TLN-**5**c	0.939	0.354	-36.8	-17.0	-19.8
	0.956±0.023	**0.310±0.038**	**-37.2±0.3**	**-16.9±0.3**	**-20.3±0.5**
TLN-**6**a	0.948	0.368	-36.7	-15.3	-21.4
TLN-**6**b	1.005	0.314	-37.1	-15.0	-22.1
TLN-**6**c	0.962	0.380	-36.6	-14.7	-21.9
	0.972±0.030	**0.354±0.035**	**-36.8±0.3**	**-15.0±0.3**	**-21.8±0.4**

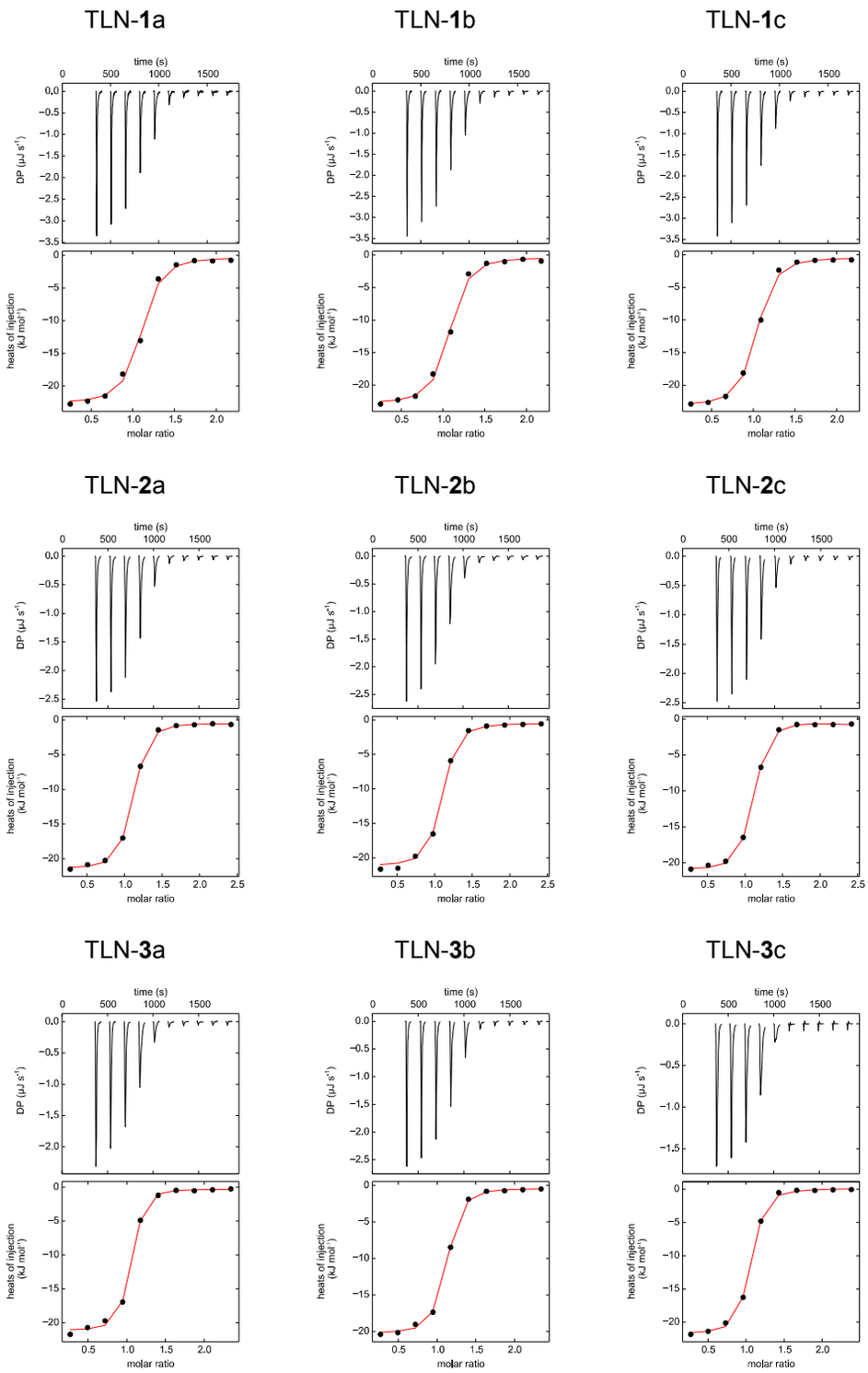


Figure S3.5. ITC measurement data of TLN-**1** to TLN-**6** each performed in triplicate (a–c). The upper panel shows the extracted peaks of the titration curve. The lower panel displays the values of the integrated peaks (*black* dots) as automatically performed by the *NITPIC* algorithm, and, fitted to them, the 1:1 model binding isotherms (*red* curve) as created with *SEDPHAT*.

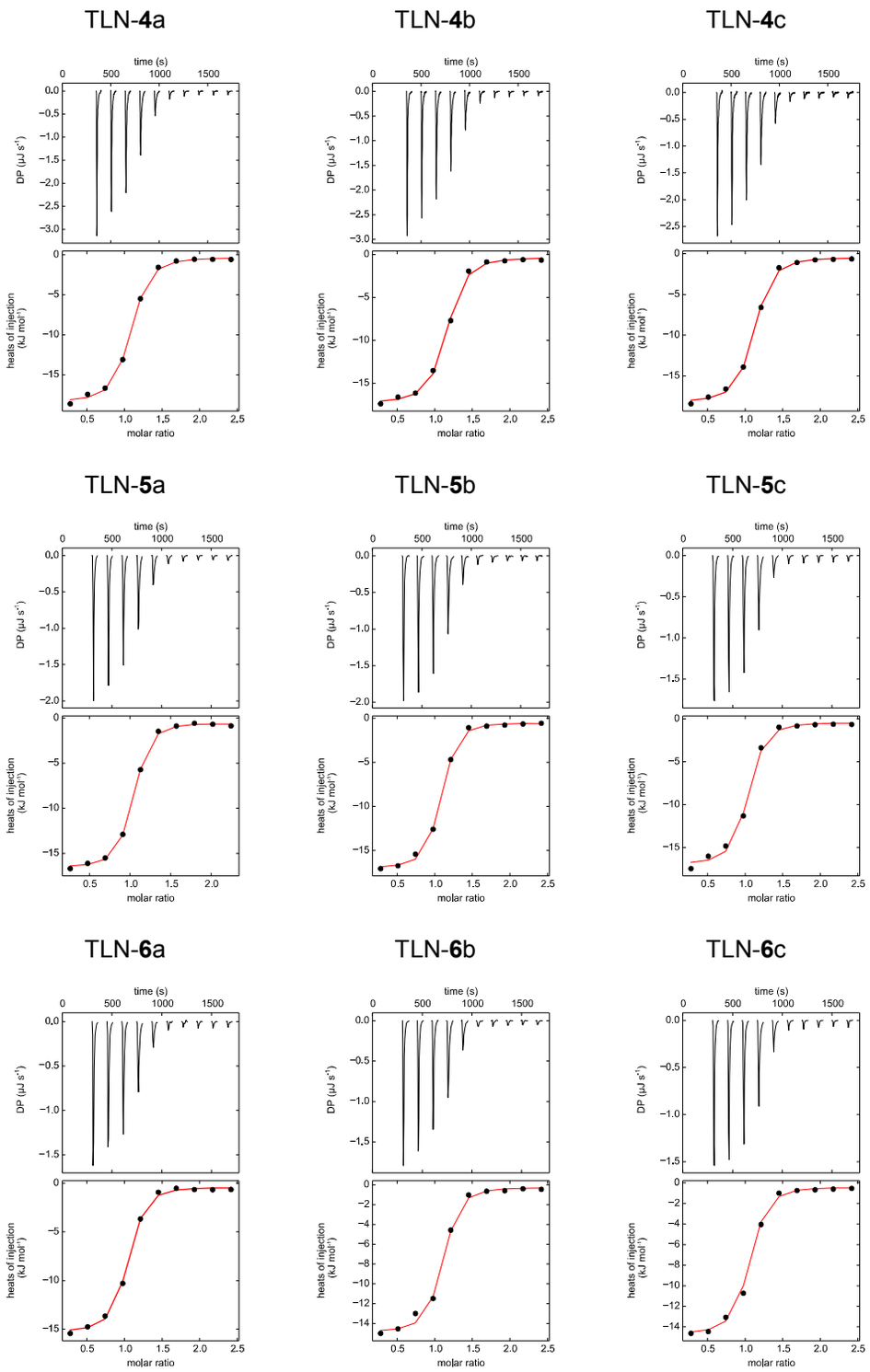


Figure S3.5. (continued)

3.10.5 ITC measurements with the addition of glycerol

For all four tested ligands (**2**, **3**, **5**, and **6**), $\Delta H°$ increases and $-T\Delta S°$ decreases with increasing concentration of glycerol (Figure S3.6). Almost complete compensation of $\Delta H°$ and $-T\Delta S°$ is observed, resulting in virtually unchanged $\Delta G°$. The relative differences of the thermodynamic parameters between the ligands remain constant. For **3** compared to **2**, identical values of $\Delta H°$ and more favorable values of $-T\Delta S°$ and $\Delta G°$ are observed for all measured glycerol concentrations.

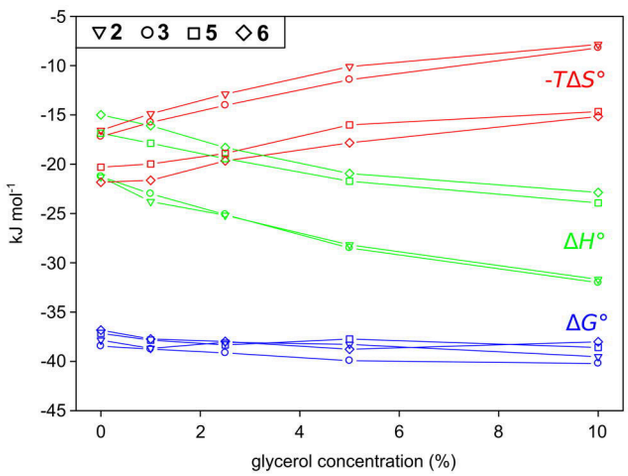


Figure S3.6. ITC measurements of **2**, **3**, **5** and **6** binding to TLN with buffers containing different glycerol concentrations of 0%, 1%, 2.5%, 5% and 10% (v/v). The symbols indicate the different ligands, the colors indicate the thermodynamic parameters $\Delta H°$ (*green*), $-T\Delta S°$ (*red*) and $\Delta G°$ (*blue*). For sake of clarity, standard deviation bars are not shown. All measurements were performed in triplicate. Data values with standard deviations are listen in Table S3.5.

Table S3.5. ITC measurement results of ligands **2**, **3**, **5** and **6** measured with the addition of different concentrations of glycerol. All measurements were performed in triplicate (a–c), from which the mean values and the standard deviations were calculated. All parameters are given as rounded numbers.

Measurement (glycerol conc.)	*n*	K_d (μM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
TLN-**2**a **(**1.0%)	1.023	0.188	-38.4	-23.8	-14.6
TLN-**2**b **(**1.0%)	1.015	0.174	-38.6	-23.6	-15.0
TLN-**2**c **(**1.0%)	1.080	0.143	-39.1	-24.0	-15.1
	1.039±0.035	**0.168±0.023**	**-38.7±0.3**	**-23.8±0.2**	**-14.9±0.3**
TLN-**2**a (2.5%)	1.087	0.188	-38.4	-25.2	-13.2
TLN-**2**b (2.5%)	1.060	0.281	-37.4	-25.1	-12.3
TLN-**2**c (2.5%)	1.034	0.186	-38.4	-25.3	-13.2
	1.060±0.027	**0.218±0.054**	**-38.1±0.6**	**-25.2±0.1**	**-12.9±0.5**
TLN-**2**a (5.0%)	1.037	0.216	-38.0	-28.5	-9.5
TLN-**2**b (5.0%)	1.014	0.186	-38.4	-28.4	-10.0
TLN-**2**c (5.0%)	1.025	0.190	-38.4	-27.6	-10.8
	1.025±0.012	**0.197±0.016**	**-38.3±0.2**	**-28.2±0.5**	**-10.1±0.6**
TLN-**2**a (10.0%)	1.023	0.146	-39.0	-31.1	-7.9
TLN-**2**b (10.0%)	1.066	0.118	-39.5	-31.4	-8.1
TLN-**2**c (10.0%)	1.104	0.098	-40.0	-32.5	-7.5
	1.064±0.041	**0.120±0.024**	**-39.5±0.5**	**-31.7±0.7**	**-7.9±0.3**
TLN-**3**a (1.0%)	1.003	0.154	-38.9	-22.7	-16.2
TLN-**3**b (1.0%)	0.981	0.158	-38.8	-22.9	-16.0
TLN-**3**c (1.0%)	0.963	0.175	-38.6	-23.4	-15.1
	0.982±0.020	**0.162±0.011**	**-38.8±0.2**	**-23.0±0.4**	**-15.8±0.5**
TLN-**3**a (2.5%)	1.012	0.165	-38.7	-24.9	-13.9
TLN-**3**b (2.5%)	0.990	0.140	-39.1	-25.5	-13.6
TLN-**3**c (2.5%)	0.991	0.116	-39.6	-25.0	-14.6
	0.998±0.012	**0.140±0.025**	**-39.1±0.4**	**-25.1±0.3**	**-14.0±0.5**
TLN-**3**a (5.0%)	0.965	0.110	-39.7	-28.8	-10.9
TLN-**3**b (5.0%)	0.962	0.101	-39.9	-28.4	-11.5
TLN-**3**c (5.0%)	0.973	0.092	-40.2	-28.3	-11.9
	0.967±0.006	**0.101±0.009**	**-39.9±0.2**	**-28.5±0.2**	**-11.4±0.5**
TLN-**3**a (10.0%)	0.976	0.106	-39.8	-31.7	-8.2
TLN-**3**b (10.0%)	0.914	0.085	-40.4	-31.7	-8.7
TLN-**3**c (10.0%)	0.927	0.081	-40.5	-32.7	-7.8
	0.939±0.033	**0.091±0.013**	**-40.2±0.3**	**-32.0±0.6**	**-8.2±0.5**
TLN-**5**a (1.0%)	1.058	0.189	-38.4	-17.9	-20.5
TLN-**5**b (1.0%)	1.015	0.289	-37.3	-17.7	-19.6
TLN-**5**c (1.0%)	1.053	0.234	-37.8	-18.0	-19.8
	1.042±0.024	**0.237±0.050**	**-37.8±0.5**	**-17.9±0.2**	**-20.0±0.4**
TLN-**5**a (2.5%)	1.042	0.195	-38.3	-19.6	-18.7
TLN-**5**b (2.5%)	1.016	0.212	-38.1	-18.9	-19.2
TLN-**5**c (2.5%)	1.054	0.171	-38.6	-19.8	-18.8
	1.037±0.019	**0.193±0.021**	**-38.3±0.3**	**-19.4±0.5**	**-18.9±0.3**
TLN-**5**a (5.0%)	0.936	0.288	-37.3	-22.1	-15.3
TLN-**5**b (5.0%)	0.933	0.210	-38.1	-21.8	-16.3
TLN-**5**c (5.0%)	0.991	0.241	-37.8	-21.3	-16.5
	0.953±0.033	**0.246±0.039**	**-37.7±0.4**	**-21.7±0.4**	**-16.0±0.7**
TLN-**5**a (10.0%)	1.067	0.152	-38.9	-24.1	-14.8
TLN-**5**b (10.0%)	1.063	0.186	-38.4	-24.1	-14.3
TLN-**5**c (10.0%)	1.042	0.185	-38.4	-23.6	-14.9
	1.057±0.013	**0.174±0.019**	**-38.6±0.3**	**-23.9±0.3**	**-14.7±0.3**
TLN-**6**a (1.0%)	0.994	0.231	-37.9	-16.2	-21.7
TLN-**6**b (1.0%)	0.986	0.270	-37.5	-16.0	-21.5
TLN-**6**c (1.0%)	0.996	0.236	-37.8	-16.1	-21.7
	0.992±0.005	**0.245±0.021**	**-37.7±0.2**	**-16.1±0.1**	**-21.6±0.2**
TLN-**6**a (2.5%)	1.020	0.212	-38.1	-18.1	-20.0
TLN-**6**b (2.5%)	0.996	0.260	-37.6	-18.7	-18.8
TLN-**6**c (2.5%)	0.981	0.198	-38.3	-18.1	-20.2
	0.999±0.020	**0.223±0.032**	**-38.0±0.4**	**-18.3±0.4**	**-19.7±0.7**
TLN-**6**a (5.0%)	0.955	0.190	-38.4	-20.9	-17.4
TLN-**6**b (5.0%)	0.939	0.142	-39.1	-21.2	-17.9
TLN-**6**c (5.0%)	0.953	0.155	-38.9	-20.7	-18.1
	0.949±0.009	**0.162±0.025**	**-38.4±0.4**	**-20.9±0.2**	**-17.8±0.4**
TLN-**6**a (10.0%)	1.065	0.199	-38.2	-23.5	-14.7
TLN-**6**b (10.0%)	1.018	0.233	-37.9	-22.2	-15.6
TLN-**6**c (10.0%)	1.012	0.226	-37.9	-22.8	-15.1
	1.032±0.018	**0.219±0.018**	**-38.0±0.2**	**-22.9±0.6**	**-15.1±0.5**

3.10.6 **ITC measurements with the addition of DMSO**

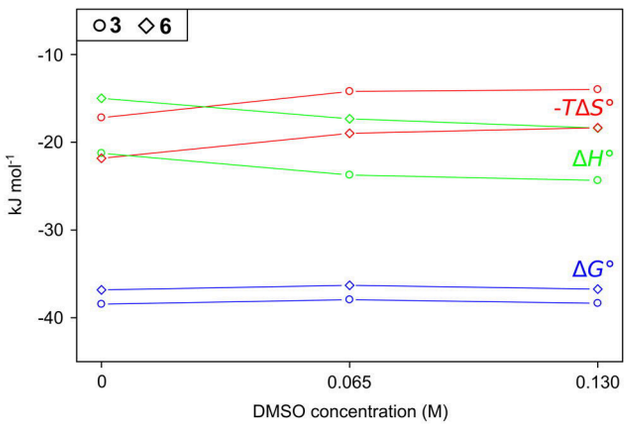


Figure S3.7. ITC measurements of **3** and **6** with the addition of DMSO. The measured DMSO concentrations of 0.065 M and 0.130 M correspond to the molar concentration of glycerol in the solutions with 5% and 10% glycerol (Figure S3.6). Data values are listed in Table S3.6.

Table S3.6. Thermodynamic parameters measured for ligands **3** and **6** by ITC with the addition of DMSO. All measurements were performed in duplicate (a+b) out of which the mean values and the standard deviations were calculated.

Measurement	*n*	K_d (μM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
TLN-**3**a (0.065 M)	0.907	0.199	-38.2	-24.3	-13.9
TLN-**3**b (0.065 M)	1.027	0.255	-37.6	-23.2	-14.5
	0.967±0.085	**0.227±0.040**	**-37.9±0.4**	**-23.7±0.8**	**-14.2±0.4**
TLN-**3**a (0.130 M)	1.014	0.208	-38.1	-24.3	-13.8
TLN-**3**b (0.130 M)	1.003	0.175	-38.6	-24.4	-14.2
	1.009±0.008	**0.192±0.024**	**-38.4±0.3**	**-24.4±0.1**	**-14.0±0.2**
TLN-**6**a (0.065 M)	1.007	0.459	-36.2	-17.2	-18.9
TLN-**6**b (0.065 M)	1.004	0.414	-36.4	-17.4	-19.0
	1.006±0.002	**0.437±0.032**	**-36.3±0.2**	**-17.3±0.1**	**-19.0±0.0**
TLN-**6**a (0.130 M)	0.992	0.3402518	-36.92	-18.36	-18.56
TLN-**6**b (0.130 M)	1.005	0.3935458	-36.56	-18.42	-18.14
	0.999±0.009	**0.367±0.038**	**-36.7±0.3**	**-18.4±0.0**	**-18.4±0.3**

3.10.7 **SPR measurements**

Table S3.7 (related to Figure 3.9). SPR measurement results of TLN-ligand complexes TLN-**1**, TLN-**2**, TLN-**3**, TLN-**5** and TLN-**6**. Kinetic analysis was performed by global analysis of single-cycle measurements performed in triplicate (Figure S3.8).

TLN complex	k_{on} (M^{-1} s^{-1})	SE (k_{on})[a]	k_{off} (s^{-1})	SE (k_{off})[a]	K_d (M)	Chi^2 (RU^2)
TLN-**1**	1.14×10^{5}	6.50×10^{2}	8.40×10^{-2}	2.70×10^{-4}	7.38×10^{-7}	0.058
TLN-**2**	7.08×10^{4}	4.10×10^{2}	5.95×10^{-2}	2.10×10^{-4}	8.41×10^{-7}	0.104
TLN-**3**	3.10×10^{4}	2.10×10^{2}	3.78×10^{-2}	1.50×10^{-4}	1.22×10^{-6}	0.257
TLN-**5**	8.58×10^{4}	7.00×10^{2}	6.53×10^{-2}	3.20×10^{-4}	7.61×10^{-7}	0.185
TLN-**6**	4.82×10^{4}	3.30×10^{2}	6.84×10^{-2}	2.70×10^{-4}	1.42×10^{-6}	0.152

[a] SE = standard error

Table S3.8 (related to Figure 3.9). Kinetic binding parameters of TLN-ligand complexes TLN-**1**, TLN-**2**, TLN-**3**, TLN-**5** and TLN-**6** as determined by individual analysis of the SPR measurements performed in triplicate, from which the mean values and the standard deviations were calculated.

TLN complex	k_{on} (M^{-1} s^{-1})	SD (k_{on})	k_{off} (s^{-1})	SD (k_{off})	K_d (M)	SD (K_d)
TLN-**1**	1.10×10^{5}	8.06×10^{3}	7.94×10^{-2}	3.97×10^{-3}	7.24×10^{-7}	8.38×10^{-8}
TLN-**2**	8.25×10^{4}	3.32×10^{4}	5.97×10^{-2}	3.07×10^{-3}	8.30×10^{-7}	4.03×10^{-7}
TLN-**3**	2.69×10^{4}	7.08×10^{3}	3.87×10^{-2}	5.09×10^{-3}	1.55×10^{-6}	5.95×10^{-7}
TLN-**5**	9.55×10^{4}	3.32×10^{4}	6.64×10^{-2}	1.49×10^{-2}	7.55×10^{-7}	2.82×10^{-7}
TLN-**6**	4.00×10^{4}	1.16×10^{4}	7.00×10^{-2}	5.98×10^{-3}	1.86×10^{-6}	5.56×10^{-7}

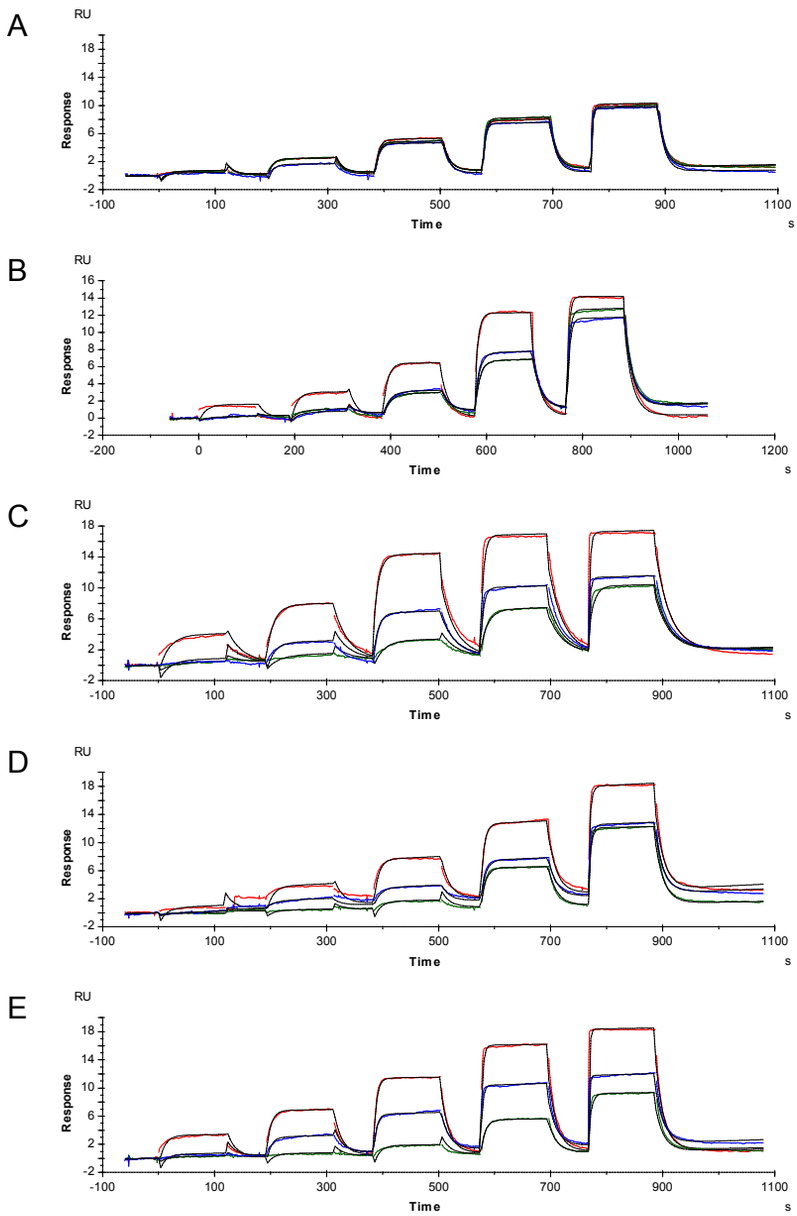


Figure S3.8 (related to Figure 3.9). SPR sensorgrams showing three separate single cycle kinetics runs in five concentrations of each **(A)** TLN-**1**, **(B)** TLN-**2**, **(C)** TLN-**3**, **(D)** TLN-**5**, and **(E)** TLN-**6**. The analyte concentration ranges spanned (A) 156–2500 nM, (B) 25–2500 nM, (C) 156–10000 nM, (D) 25–2500 nM, (E)156–10000 nM. The colouring of sensorgrams represents the concentrations used in each cycle (A) 156, 312, 625, 1250, 2500 nM (*red*, *blue*, *green*), (B) and (D) 25–2500 nM (156, 312, 625, 1250, 2500 nM (*red*), 25, 74, 222, 667, 2000 nM (*blue*, *green*)), (C) and (E) 156, 312, 625, 1250, 2500 nM (*blue*, *green*), 625, 1250, 2500, 5000, 10000 nM (*red*). Global analysis of the triplicate data was performed (*black*) to account for experimental differences between runs such as variation in capture level (800–1200 RU), drift and used analyte concentrations. The concentration range used spanned

from <0.2×K_d to >3.0×K_d for all compounds (K_d values determined by SPR are presented in Table S3.7).

3.10.8 **Calculation of buried SASAs**

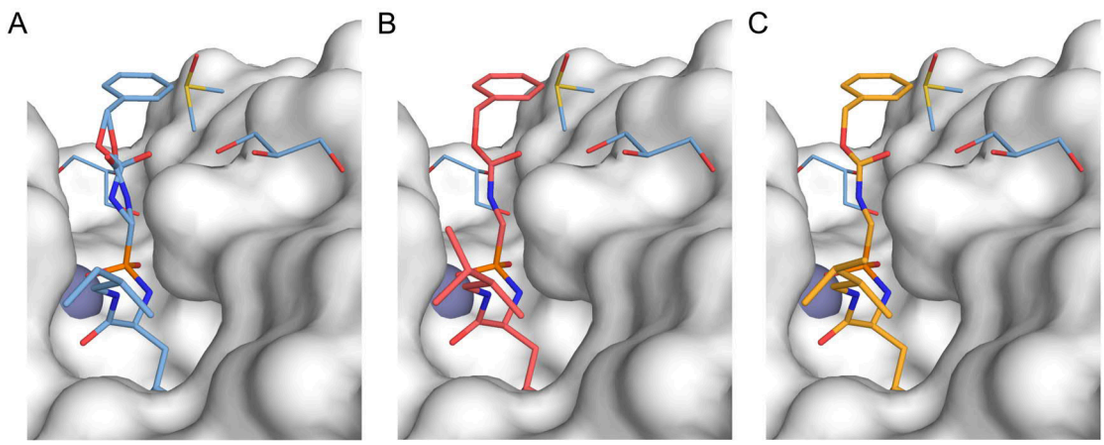


Figure S3.9. Crystal structure TLN-**5** with the missing terminal methyl group of the P_2' portion (panel A, bound ligand in *blue*), and the two modeled conformations **5**A_m (panel B, *red*) and **5**B_m (panel C, *yellow*) of the missing methyl group used for the calculations of the SASAs.

Table S3.9 (related to Figure 3.10). Calculated total and buried solvent accessible surface areas (SASAs) of ligands **1**–**6** in complex with TLN.

	Ligand	SASA (Å^2)	
		total	buried
1		702	512
2	conformation *A*	726	536
	conformation *B*	726	535
3		740	548
4		705	566
5	missing P_2' methyl group (Figure S3.9A)	704	520
	methyl group modeled in conformation A_m (Figure S3.9B)	729	539
	methyl group modeled in conformation B_m (Figure S3.9C)	729	539
6		748	555

3.10.9 **Comparison MD simulations and crystal structures of TLN-2 and TLN-5**

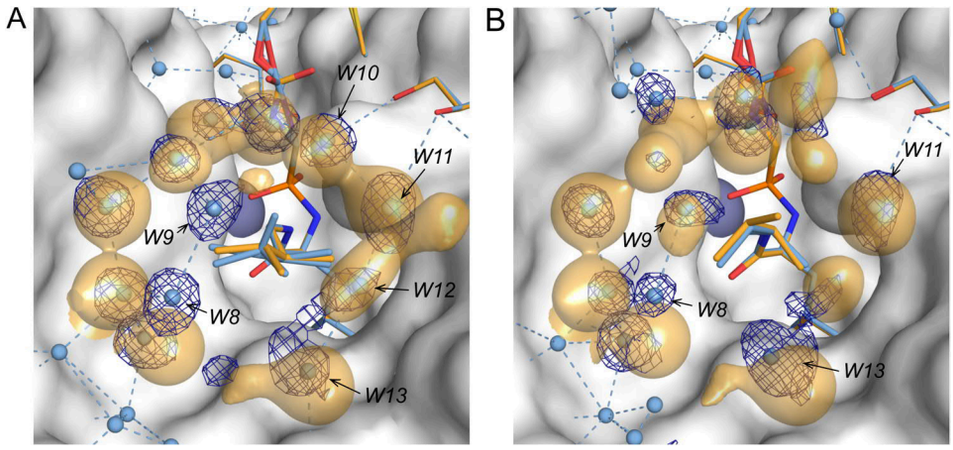


Figure S3.10 (related to Figure 3.11). Prediction of positions of water molecules as calculated by MD simulations in comparison to the crystallographically observed electron densities and refined crystal structure models of **(A)** TLN-**2** and **(B)** TLN-**5**. The modeled coordinates of ligand, glycerol and DMSO molecules used in the MD simulations are displayed as *yellow* stick models with color-coded heteroatoms. The *yellow*, semitransparently contoured regions show computed areas in the first solvation layer of the P_2' groups with an occupancy probability by a water molecule of at least 48%. The crystallographically determined binding modes of ligand and additive molecules are superimposed as *blue* stick models with color-coded heteroatoms. Water molecule positions determined in the crystal structures are displayed as *blue* spheres, and the F_o–F_c omit electron density is displayed as *dark blue* mesh at a contour level of 3σ for the water molecules positioned in the first solvation layer of the P_2' groups. H-bond distances are indicated as *blue* dotted lines. Positions of water molecules, which are discussed in the main text, are labeled with identifiers according to Figure 3.4. The solvent excluded surface of TLN is shown in *white*.

Chapter 4

Paying the Price of Desolvation in Solvent-Exposed Protein Pockets: Impact of Solubilizing Polar Groups on Affinity and Binding Thermodynamics in a Series of Thermolysin Inhibitors

Jonathan Cramer,[1] Stefan G. Krimmer,[1]
Andreas Heine,[1] Gerhard Klebe[1]

[1]Department of Pharmaceutical Chemistry, University of Marburg,
Marbacher Weg 6, 35032 Marburg, Germany

Manuscript submitted

4.1 Abstract

In lead optimization programs in medicinal chemistry, open, solvent-exposed protein pockets are often disregarded as prospective binding sites. Due the proximity of the bulk solvent, many researchers are instead enticed to attach charged and polar groups to a respective position of an inhibitor scaffold in order to improve solubility and pharmacokinetic properties of a lead candidate. It is rarely considered that solvent effects from the organization of water molecules in the first hydration shell of a protein–ligand complex can have a significant impact on the thermodynamics of binding. In this case study, we aimed to investigate the thermodynamic fingerprint of a series of thermolysin inhibitors featuring a charged ammonium group that is gradually pulled further into the solvent-exposed S_2' pocket of the protease. It was found that the cost for the partial desolvation of the polar group at the solvent interface could not easily be compensated by interactions to the protein or surrounding water molecules. In a direct comparison to hydrophobic analogues, a significant loss of affinity by a factor of up to 180 in K_d was recorded.

Keywords

protein–ligand interactions, binding thermodynamics, hydrophobic effect, desolvation, X-ray crystallography, isothermal titration calorimetry

4.2 Introduction

During the course of a protein–ligand binding event, the solvation shell of both partners plays an active role for the recognition process. Water molecules interact with both species in their solvated state in bulk solution prior to binding. Thus, before a binding event is initiated, the hydration shell of either molecule has to be partially shed and adapted to allow a direct interaction between the protein and its ligand. This process is known as desolvation [173,174]. The subsequently formed protein–ligand complex exposes a newly generated surface that again necessitates an adjustment of the surrounding solvation layer. It is therefore appropriate to consider water molecules as a third binding partner in biological recognition events. The thermodynamic signature associated with the reorganization of solvent molecules can be diverse. Most commonly, the desolvation of hydrophobic ligand portions is believed to result in an entropically driven free energy gain. This so called "hydrophobic effect" is said to originate from the displacement of conformationally restricted water molecules from an apolar interface into the bulk [7,175]. However, it has been shown in multiple studies that this simple model does not hold true in all cases [9,10,15,16,93,124,176,177]. Depending on the characteristics of the interface, most notably the shape of the respective surface[14,177], the partitioning of energy contributions from the displacement of water molecules can be diverse and driven by enthalpy, as well as entropy. On the other end of the spectrum, polar components of solutes actively engage in hydrophilic interactions to the water molecules comprising their hydration shell. The desolvation of a polar moiety is commonly accompanied by an enthalpic penalty that arises from the loss of a hydrophilic contact, such as hydrogen bonds, to the solvent molecules in its hydration shell [173]. In order to compensate for that energetic loss, an equivalent interaction to a binding partner at the protein binding site has to be established to prevent a free energy deficit [113,120].

In lead optimization programs in medicinal chemistry, charged and polar functional groups are commonly included in an inhibitor scaffold as means to improve solubility or pharmacokinetic properties of a prospective drug candidate. If the targeted protein does not provide a specific interaction for these groups, it is often attempted to place polar components in open, solvent accessible areas, where they interfere with the solvent interface of the protein–ligand complex. The underlying assumption of this strategy is that a partial desolvation of the polar group does not impose a significant penalty to the binding affinity. A systematic study of the thermodynamics that originate from solvation effects of polar groups at protein–ligand interfaces, however, has not been realized yet.

In previous studies, we have used the interaction between the metalloprotease thermolysin (TLN) and carefully tailored phosphonopeptide inhibitors as a model system for the investigation of solvent effects during protein–ligand binding. These studies focused on the rearrangement of water molecules around a newly generated hydrophobic surface around the

solvent accessible, mostly apolar S_2' pocket of the protein. For the research presented in this contribution, we envisaged an inhibitor system that would be able to introduce a terminal, increasingly exposed charged and polar functional group into the surface generated by the formed protein–ligand complex. The hydrophilic moiety was supposed to act as an anchor point for water molecules in the hydration shell and introduce an augmented ordering of the solvent network. By an incremental retraction of the polar group deeper into the hydrophobic pocket, we hoped to be able to study the thermodynamic characteristics of the process of desolvation and gain insight that may prove useful for the development of new drugs, with a particular focus on typical protocols to improve pharmacokinetics.

As a suitable inhibitor system for the abovementioned research effort, two series, each consisting of four compounds, were synthesized. While the basic phosphonopeptide scaffold of TLN inhibitors was kept constant, a primary amino group, which would be charged under the assay conditions, was introduced in the P_2' position of the ligands (Figure 4.1). For inhibitors **1**–**4**, an aminoalkyl chain was included in this position, whereas **5**–**8** contained an amino acid scaffold with an equivalent amino function in the side chain. The thermodynamic profiles of the interaction of inhibitors **1**–**8** were determined by isothermal titration calorimetry (ITC) and the binding modes were analyzed by protein X-ray crystallography.

4.3 Results

4.3.1 ITC data collection.

As shown previously, the binding of phosphonamidate type inhibitors to TLN is accompanied by the release of a proton from the buffer [93]. During the transition from the open to the closed, ligand-bound state, Glu143 is believed to change its protonation state as it recedes from the partly exposed binding site into the S1 pocket. The calorimetric experiment records this protonation linkage in form of a buffer dependent contribution, superimposed on the enthalpy of binding resulting from the protein–ligand interaction. For an accurate assessment of the net thermodynamic data of binding, the recorded enthalpy values have to be corrected for this superimposed enthalpy of ionization.

Therefore, the thermodynamic binding profiles of **1**–**8** were determined by ITC in three different buffers of deviating ionization enthalpy. A fit of the recorded enthalpy against the reported enthalpy of ionization to a linear model gave the buffer corrected enthalpy, as well as information about the number of protons that are transferred during ligand binding (Figure 4.1 and Figure S4.4) [96]. This value was determined to fall into a range between 0.6 and 0.9 moles of protons for these ligands. Considering the total error across all performed

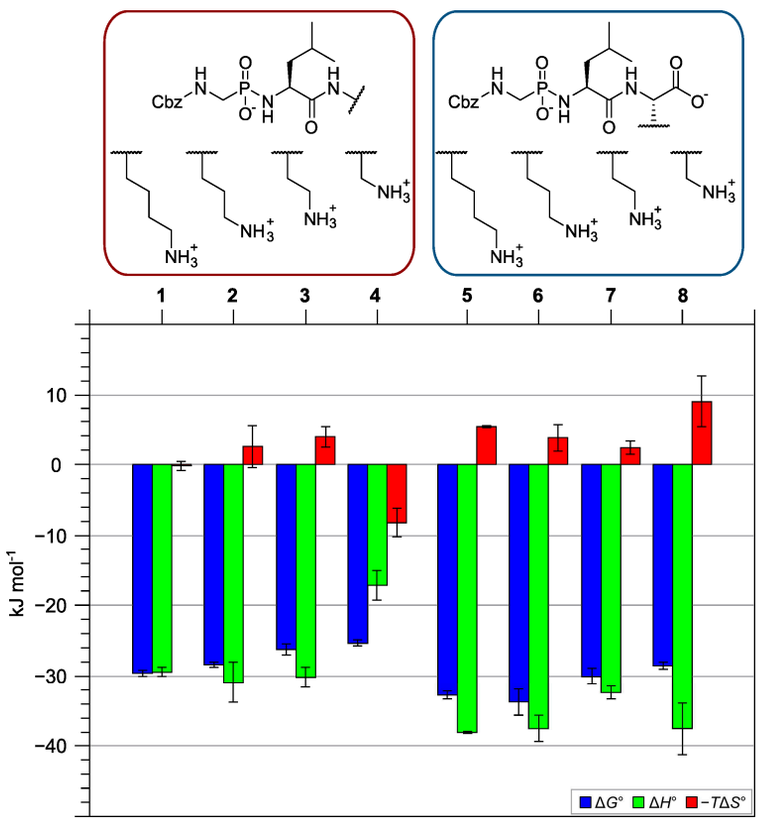


Figure 4.1. Thermodynamic data from ITC experiments. Gibb's free energy (Δ*G*°, blue), enthalpy (Δ*H*°, green), and entropy of binding (–*T*Δ*S*°, red).

experiments, we conclude that during the binding of **1**–**8**, a roughly equivalent amount of protons is picked up by the buffer. A statistical test for the equality of the regression lines did not identify a significant deviation within the sample set [178].

Figure 6.1 gives an overview over the thermodynamic parameters of ligand binding after a correction for the buffer ionization contribution. Within the aliphatic amine series (**1**–**4**), the free energy of binding shows a steady decline with a reduction of the chain length. From **1**–**3**, this effect is accompanied by a moderate entropic loss ($-T\Delta\Delta S°_{1\rightarrow2}$= 2.7 kJ mol^{-1}; $-T\Delta\Delta S°_{1\rightarrow3}$= 4.1 kJ mol^{-1}) that is only partly compensated by an enthalpic gain ($\Delta\Delta H°_{1\rightarrow2}$= –1.5 kJ mol^{-1}; $\Delta\Delta H°_{1\rightarrow3}$= –0.7 kJ mol^{-1}). A further contraction of the aliphatic chain from **3** to **4**, however, reveals a drastic reversal of the thermodynamic contributions. The enthalpy of binding is reduced significantly ($\Delta\Delta H°_{3\rightarrow4}$= 13.0 kJ mol^{-1}), while a pronounced entropic benefit ($-T\Delta\Delta S°_{3\rightarrow4}$= –12.1 kJ mol^{-1}) compensates for that loss. Within the amino acid series (**5**–**8**), a slight increase in Δ*G*° from **5**–**6** ($\Delta\Delta G°_{5\rightarrow6}$= –1.0 kJ mol^{-1}), followed by a steady affinity loss upon reduction of the chain length is observed ($\Delta\Delta G°_{6\rightarrow7}$= 3.7 kJ mol^{-1}; $\Delta\Delta G°_{7\rightarrow8}$= 1.9 kJ mol^{-1}). From **5**–**7**, the change in binding free energy is partitioned into an enthalpic loss ($\Delta\Delta H°_{5\rightarrow6}$=

0.5 kJ mol^{-1}; $\Delta\Delta H°_{6\rightarrow7}$= 5.1 kJ mol^{-1}) and a compensating favorable entropic contribution (–$T\Delta\Delta S°_{5\rightarrow6}$= –1.5 kJ mol^{-1}; –$T\Delta\Delta S°_{6\rightarrow7}$= –1.4 kJ mol^{-1}). For **8**, the inhibitor with the shortest chain length of this series, a significant loss in entropy (–$T\Delta\Delta S°_{7\rightarrow8}$= 6.6 kJ mol^{-1}) with a counteracting enthalpic gain ($\Delta\Delta H°_{7\rightarrow8}$= 5.1 kJ mol^{-1}) is recorded. In comparison to their polar analogues, the ligands featuring a hydrophobic substituent in the P2' position display a significantly increased affinity, evident by a gain in free energy of binding ($\Delta\Delta G°_{3\rightarrow9}$ = 9.9 kJ mol^{-1}, $\Delta\Delta G°_{4\rightarrow10}$ = 7.5 kJ mol^{-1}, $\Delta\Delta G°_{7\rightarrow11}$ = 10.1 kJ mol^{-1}, $\Delta\Delta G°_{8\rightarrow12}$ = 12.1 kJ mol^{-1}). This effect is mainly driven by a dramatic loss in enthalpy for the polar inhibitors ($\Delta\Delta H°_{3\rightarrow9}$ = 11.0 kJ mol^{-1}, $\Delta\Delta H°_{4\rightarrow10}$ = 16.0 kJ mol^{-1}, $\Delta\Delta H°_{7\rightarrow11}$ = 18.8 kJ mol^{-1}, $\Delta\Delta H°_{8\rightarrow12}$ = 16.5 kJ mol^{-1}).

4.3.2 X-ray crystallography

The crystal structures of **1**–**8** (Figure 4.2) reveal that their overall binding mode to TLN is highly conserved across the inhibitor series. Significant differences are only observed in the chemically deviating P_2' moieties and the surrounding hydration pattern. In TLN-**1**, the terminal amino group of ligand **1** is hydrogen-bonded to the backbone carbonyl oxygen of Asn111. No additional water molecules could be confidently assigned to the electron density within interaction distance to the ligand. Residual electron density peaks in the mF_o–DF_c map, however, suggest the presence of weakly populated, transient water sites (Figure S4.5). In the structure of **2**, the truncated alkyl chain still allows the formation of a hydrogen bond to Asn111. The deviating geometry of this interaction, however, supports the stabilization of two water molecules in the hydration shell of the terminal ammonium group. A further reduction of the chain length in TLN-**3** prevents the formation of a hydrogen bond to Asn111. Instead, a water molecule is fixed at a position that was formerly occupied by the amino groups of **1** and **2**. The terminal aminoethyl chain of **3** is oriented toward the side, where a conserved water cluster at the rim of the S_2' pocket is found. It engages in two contacts to water molecules. The aminomethyl group of **4**, the shortest ligand of the series, immerges even further into the hydrophobic S_2' pocket. In the crystal structure, its side chain is considerably disordered, so that two conformations had to be considered in the refinement model to explain the observed electron density. In one conformation (26% occupancy), the amino group is oriented toward the upper part of the pocket, where it is able to form hydrogen bonds to two conserved water molecules. In the second conformation (74% occupancy), it is rotated toward the hydrophobic bottom of the pocket. This allows the charged nitrogen to interact with a conserved water molecule on the left hand side of the binding pocket. In addition, the water molecule interacting with Asn111 in **3**, is dragged into the hydrophobic crevice, where it acts as a second interaction partner for one conformer of the aminomethyl group of **4**. This arrangement

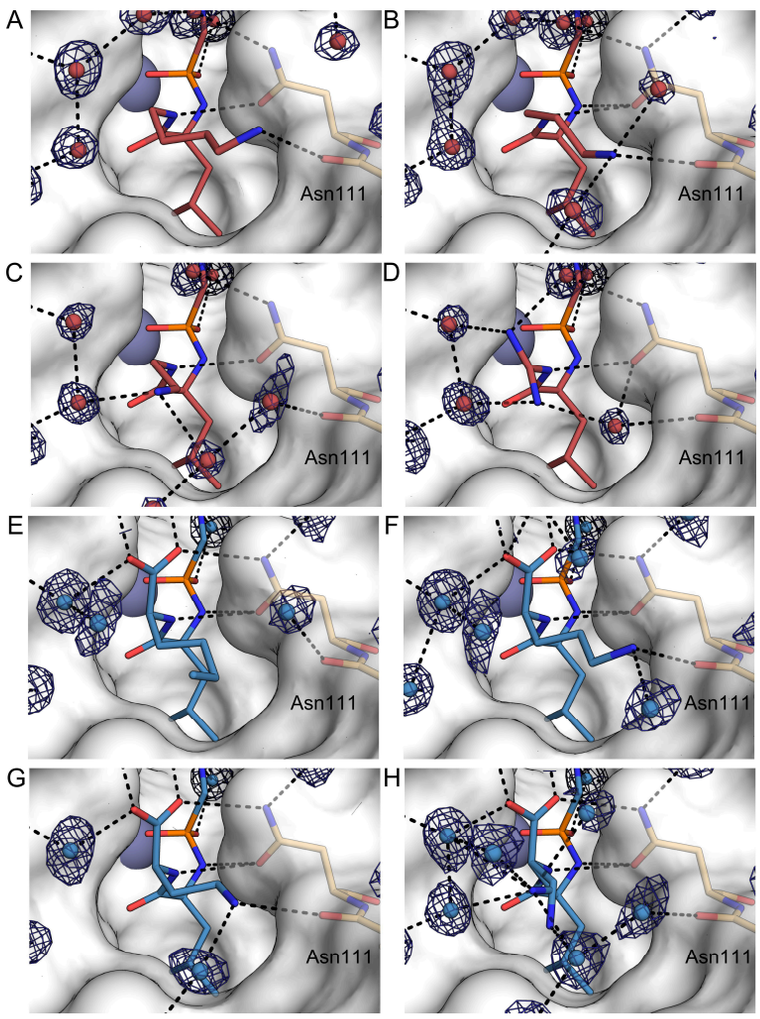


Figure 4.2. Binding modes of TLN inhibitors **1** (A), **2** (B), **3** (C), **4** (D), **5** (E), **6** (F), **7** (G), and **8** (H). Carbon atoms of the inhibitors are shown in red (**1**–**4**) or blue (**5**–**8**), protein residues are displayed in beige. Heteroatoms are color coded by atom type. The white surface represents the solvent excluded protein surface. Hydrogen bonds are indicated as dashed lines. The 2mF_o–DF_c density map for the water molecules is shown as a blue mesh (1.0σ).

extends the distance between the interaction partners from 2.8 Å (in TLN-**3**) to 3.0 Å, thereby weakening the interaction to the protein (based on hydrogen bond distance criteria).

Inhibitors **5**–**8** feature a carboxylic acid group at their P_2' moiety in addition to the aminoalkyl side chain. In all structures, the carboxy function is fixed in position by a strong interaction to Asn112. This interaction is believed to increase the affinity of a compound by interfering with the release mechanism during the dissociation process of the bound ligands [179]. In addition, the terminal carboxy function of **5**–**8** engages in strong, charge-assisted hydrogen bonds to conserved water molecules that form energetically favorable polygonal networks above the

pocket. The lysine derivative **5** binds in a position, in which the terminal ammonium group sticks out into the solvent. Due to its high residual mobility, it is scattered over multiple positions and thus, not clearly defined in the electron density. The length of the side chain of this ligand does not allow an interaction with Asn111. Instead, a water molecule donates a hydrogen bond to this residue. The shorter chain length of the ornithine derivative **6**, however, places the ammonium side chain within an optimal interaction distance to Asn111, while simultaneously keeping the orientation of the carboxy group unchanged. A water molecule, placed toward the front of the pocket, provides an additional interaction. The binding mode of the homolog **7** is very similar to **6**. The shorter chain length in **7**, however, does not allow an optimal interaction geometry of the hydrogen bond to Asn111. While the strong interactions of the carboxy group keep its position fixed, the truncation of the side chain increases the hydrogen bond distance to Asn111 from 2.8 Å (in TLN-**6**) to 3.0 Å in TLN-**7**, thereby weakening the interaction. Nonetheless, the placement of the terminal ammonium group allows the recruitment of water molecules that bind toward the lower rim of the S_2' pocket. The crystal structure of **8** reveals that the distance of the terminal ammonium group of this compound to Asn111 is too large for a hydrogen bond interaction. As a result, the charged functionality is disordered over multiple orientations (Figure S4.6), nevertheless interacting quite extensively with the surrounding hydration shell. In total, six water molecules are trapped in the close neighborhood of the S_2' pocket and fixed in positions, wrapping around the side chain of **8**. One of these waters takes up the interaction to Asn111.

4.4 Discussion

4.4.1 Thermodynamics–Structure Relationship for 1–4

The thermodynamic binding profile of **2** in comparison to **1** is characterized by an increase in $\Delta G°$ (lower affinity). This can be attributed to an enthalpic gain that is overcompensated by a less beneficial entropic signal. Both ligands show a strong interaction with Asn111 (hydrogen bond length 2.8 Å). The more favorable enthalpy of **2** might originate from a stronger fixation of its aminoalkyl chain. In comparison to **1**, **2** seems to provide a better environment for the stabilization of an additional water molecule in the hydration shell of the ligand, which in turn reduces the residual flexibility of the aliphatic chain. The stronger fixation of **2** strengthens the enthalpic interactions while the entropic contribution is reduced. Considering the experimental error for the determination of enthalpic and entropic contributions, this effect is not very distinctive. From **2** to **3**, the overall decrease in affinity is driven by a slightly less favorable entropic contribution in combination with an increase in enthalpy. Again, the change in enthalpy and entropy is not very pronounced and lies within the standard deviation of the measurement. Due to an increased distance to Asn111, the hydrogen bond to this residue

Table 4.1. Data collection and refinement statistics for crystal structures TLN-**1** – TLN-**4**.

	Complex (PDB code)			
	TLN-**1** (5N2Z)	TLN-**2** (5N2T)	TLN-**3** (5N34)	TLN-**4** (5N3V)
(A) Data collection and processing				
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters: *a*, *b*, *c* (Å)	93.0, 93.0, 130.9	93.1, 93.1, 130.8	92.5, 92.5, 129.9	92.7, 92.7, 130.1
Matthews coefficient (Å^3/Da)[a]	2.4	2.4	2.4	2.4
Solvent content (%)[a]	48	48	48	48
(B) Diffraction data				
Resolution range (Å)	50–1.37 (1.45–1.37)	50–1.38 (1.46–1.38)	50–1.22 (1.29–1.22)	50–1.12 (1.19–1.12)
Unique reflections	70483 (11165)	69141 (10996)	94021(14606)	125558 (19848)
$R(I)$sym (%)	5.8 (48.6)	6.6 (47.7)	8.5 (48.8)	8.8 (45.5)
Wilson *B* factor (Å^2)	11.0	11.5	8.4	8.5
Completeness (%)	99.9 (99.5)	99.5 (99.8)	96.3 (94.0)	99.4 (99.6)
Redundancy	11.9 (12.1)	11.9 (12.0)	5.7 (6.0)	5.5 (5.4)
$\langle I/\sigma(I)\rangle$	30.0 (5.2)	28.4 (4.9)	13.3 (3.7)	11.5 (3.2)
(C) Refinement				
Resolution range (Å)	46.51–1.37	46.53–1.38	30.10–1.22	40.13–1.12
Reflections used in refinement (work/free)	70477/3525	69136/3457	95006/4700	125548/6278
Final *R* value for all reflections (work/free) (%)	11.1/13.5	11.2/13.8	11.0/13.0	11.0/12.8
Protein residues	316	316	316	316
Calcium/zinc ions	4/1	4/1	4/1	4/1
Inhibitor atoms	29	28	27	55
Water molecules	376	369	339	372
RMSD from ideality:				
Bond lengths (Å)	0.011	0.007	0.012	0.010
Bond angles (°)	1.1	1.0	1.2	1.1
Ramachandran plot:[b]				
Residues in most favored regions (%)	88.1	88.5	88.1	88.5
Residues in additionally allowed regions (%)	10.7	10.4	11.1	10.4
Residues in generously allowed regions (%)	0.7	0.7	0.4	0.7
Residues in disallowed regions (%)[c]	0.4	0.4	0.4	0.4
Mean *B* factors (Å^2):[d]				
Protein side chain atoms	12.2	12.6	9.8	10.2
Protein main chain atoms	10.9	11.4	8.7	9.0
Inhibitor	12.3	12.0	9.2	8.4
Water molecules	26.4	26.3	23.4	23.9

[a] Matthews coefficients and solvent contents were calculated with the program *Matthews_coef* from the *CCP4 suite*.[180] [b] Ramachandran plots were calculated with *PROCHECK*.[138] [c] The Ramachandran outlier is Thr26 as described in literature.[181] [d] Mean *B* factors were calculated with *MOLEMAN*.[182]

Table 4.2. Data collection and refinement statistics for crystal structures TLN-**5** – TLN-**8**.

	Complex (PDB code)			
	TLN-**5** (5N31)	TLN-**6** (5N2X)	TLN-**7** (5MNR)	TLN-**8** (5N3Y)
(A) Data collection and processing				
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters: *a*, *b*, *c* (Å)	92.8, 92.8, 130.9	92.7, 92.7, 130.3	92.7, 92.7, 130.5	92.7, 92.7, 130.2
Matthews coefficient (Å^3/Da)[a]	2.4	2.4	2.4	2.4
Solvent content (%)[a]	48	48	48	48
(B) Diffraction data				
Resolution range (Å)	50–1.37 (1.45–1.37)	50–1.21 (1.28–1.21)	50–1.25 (1.32–1.25)	50–1.34 (1.42–1.34)
Unique reflections	70390 (11100)	100719 (15957)	91757(14579)	74695 (11813)
$R(I)$sym (%)	6.8 (47.8)	6.7 (49.1)	6.6 (47.8)	6.6 (48.1)
Wilson *B* factor (Å^2)	10.4	9.4	9.3	10.2
Completeness (%)	99.8 (99.1)	99.8 (99.3)	99.8 (99.5)	99.9 (99.3)
Redundancy	9.6 (9.3)	11.8 (12.1)	12.8 (12.9)	9.6 (9.8)
$<I/\sigma(I)>$	23.0 (4.6)	23.9 (4.7)	25.4 (4.9)	22.6 (4.8)
(C) Refinement				
Resolution range (Å)	26.74–1.37	40.16–1.21	40.16–1.25	40.13–1.34
Reflections used in refinement (work/free)	70383/3519	100710/5036	91748/4588	74691/3735
Final *R* value for all reflections (work/free) (%)	10.7/13.4	11.1/12.7	10.3/12.5	11.0/13.8
Protein residues	316	316	316	316
Calcium/zinc ions	4/1	4/1	4/1	4/1
Inhibitor atoms	33	31	31	32
Water molecules	401	361	413	372
RMSD from ideality:				
Bond lengths (Å)	0.009	0.007	0.011	0.008
Bond angles (°)	1.1	1.0	1.2	1.0
Ramachandran plot:[b]				
Residues in most favored regions (%)	88.5	88.1	88.5	88.1
Residues in additionally allowed regions (%)	10.4	10.7	10.4	10.7
Residues in generously allowed regions (%)	0.7	0.7	0.7	0.7
Residues in disallowed regions (%)[c]	0.4	0.4	0.4	0.4
Mean *B* factors (Å^2):[d]				
Protein side chain atoms	11.4	10.9	10.7	11.0
Protein main chain atoms	10.2	9.7	9.5	9.8
Inhibitor	11.2	9.3	9.8	9.3
Water molecules	26.8	22.8	25.4	24.3

[a] Matthews coefficients and solvent contents were calculated with the program *Matthews_coef* from the *CCP4 suite*.[180] [b] Ramachandran plots were calculated with *PROCHECK*.[138] [c] The Ramachandran outlier is Thr26 as described in literature.[181] [d] Mean *B* factors were calculated with *MOLEMAN*.[182]

is no longer maintained in TLN-**3**. Instead, the charged ammonium group is fixed in position by an interaction to two water molecules, which mediate the contact to Asn111. A further reduction of the chain length introduces a significant shift in the thermodynamics of binding. For **4**, a rather dramatic enthalpic penalty is recorded. This is partly compensated by a gain in entropy, resulting in a reduced overall affinity. An explanation for this finding might be the observation that the charged and polar ammonium moiety in **4** is pulled back even further into the more hydrophobic part of the S_2' pocket. In addition, the terminal ammonium group is scattered over two orientations, which parallels with an entropic advantage. A water molecule that engages in the conserved interaction to Asn111 is forced to adopt a position deep within the hydrophobic pocket in order to maintain a hydrogen bond to both interaction partners. These manifold changes compared to other complexes of the series result in an unfavorable situation, which is reflected in the enthalpic penalty of $\Delta\Delta H°_{3\rightarrow4}$= 13.0 kJ mol^{-1}. The compensating entropic benefit of $-T\Delta\Delta S°_{3\rightarrow4}$= −12.1 kJ mol^{-1} may originate from an increased mobility of the water molecules in the surrounding hydration shell, along with the apparent disorder of the ligand side chain, as evident from the two refined conformations.

4.4.2 Thermodynamics–Structure Relationship for 5–8

As reported for a related series of inhibitors,[94] the addition of a carboxy group in **5**–**8** imposes a strong ordering of the water structure around the S_2' pocket. For all ligands, the carboxylic acid group is tightly fixed in position by Asn112 and interacts with a cluster of structurally conserved water molecules. In comparison to **1**–**4**, this arrangement leads to a gain in affinity that is dominated by a favorable enthalpic contribution. The side chain of the lysine derivative **5** sticks out into the solvent and the terminal ammonium moiety is disordered in the crystal structure. The residual solvation pattern of ordered water molecules toward the lower rim of the binding pocket appears rather imperfect. A reduction of the chain length to **6** is accompanied by a slight decrease in the entropy of binding. In TLN-**6**, the ornithine side chain is perfectly positioned for an interaction to Asn111 and adopts an orientation very similar to that observed with the isostructural side chain of **1**. A reason for the minor alterations in the thermodynamic profile from **5** to **6** might be the displacement of a water molecule in TLN-**5** with the ammonium side chain of **6**. Conceivably, the loss in mobility of the aliphatic chain is overcompensated by the gain of mobility of the water molecule that is displaced from the complex. Regarding the enthalpic term, the polar interactions of the terminal ammonium group of the lysine side chain in **5**, which only interacts with the hydrating water molecules in the bulk phase, are compensated by the hydrogen bond interaction of **6** with Asn111 and an additional water molecule that is entrapped by the complex. However, this effect is not very distinctively alternating the overall thermodynamic signature.. In the structure of **7**, it becomes apparent that the strong polar interaction of the carboxy terminus dominates the binding

mode of the P_2' residue. This group remains fixed at its position in contact with Asn112, while the truncation of the side chain deteriorates the interaction to Asn111. Due to the shorter chain length of the diaminobutyric acid in **7**, the hydrogen bond distance to this residue is increased to 3.0 Å, thus weakening the interaction. This correlates with a loss in enthalpy in the thermodynamic binding profile. A further reduction of the side chain length to the diaminopropionic acid derivative **8** pulls the polar ammonium moiety further back toward the hydrophobic part of the binding pocket. In the crystal structure, the aminomethyl side chain of **8** is disordered over two conformations. The electron density even provides evidence for a third conformation, which could not be included in the model with sufficient confidence (Figure S4.6). Similar properties of the side chain are not evident in the crystal structure of the related complex TLN-**3**. A total of six water molecules are stabilized in the first hydration layer around the S_2' pocket. They are engaged in a pronounced network of hydrogen-bond interactions to the polar side chain of **8**, as well as Asn111. In the thermodynamic profile, the binding event is characterized by an enthalpic advantage of **8** over **7**. At the same time, an unfavorable entropic contribution is recorded. The recorded gain in enthalpy might originate from the strong polar interactions within the pronounced water network entrapped in the complex with **8**. The stabilization of a similar solvent network above the protein–ligand complex has been shown to result in an enthalpically favorable contribution to binding.[16] The fixation of this large amount of solvent molecules is also reflected in the binding profile as a pronounced entropic disadvantage. However, the multiple conformations of the aminomethyl side chain parallel an increased residual mobility of the ligand, which likely counteracts this effect. In total, a trend toward decreasing affinity upon reduction of the chain length bearing the charged group can be observed in both series, while the partitioning in enthalpic and entropic contributions vary between the ligands as a result of complex specific features originating from differences in the complexity of the established hydration pattern, residual ligand mobility and inventory of formed hydrogen bonds. With respect to the adopted conformation of the side chain, the isostructural pairs **1**/**6**, **2**/**7**, and **3**/**8** show similar features.

4.4.3 Discussion of thermodynamic affinity data

Within the two investigated ligand series, the overall affinity shows a steady decrease with a reduction of the chain length. While this was an expected outcome of the experiment, the effect of trapping a charged and polar ammonium group increasingly closer to a hydrophobic environment turned out to be less dramatic than we initially assumed. For **1**–**4**, the difference in $\Delta G°$ between the most and least potent inhibitors (**1** and **4**) equals 4.2 kJ mol^{-1}, which corresponds to a factor of 6 in K_d. The potency of the carboxy ligands **5**–**8** differs by 5.2 kJ mol^{-1} in $\Delta G°$ (**6** and **8**), which equals a factor of 7 in K_d. In contrast to this, the variation of a purely hydrophobic chain in the P_2' position of related peptidomimetic TLN inhibitors, which were

previously studied by us, resulted in a difference in K_d by a factor of 50.[15] In addition to the burial of hydrophobic surface area, the local ordering of water molecules in thermodynamically beneficial networks proved to be important for the affinity gain in the latter cases.[15,16] A comparison of the water structure around the S_2' pocket of TLN-**1** – TLN-**8** to the structures of related hydrophobic analogues reveals that hardly any additional water molecules that are observable in the solid state by crystallography are recruited by the complexes of the polar compounds. Remarkably, in many cases the apolar and aliphatic substituents stabilize a more elaborate network of solvent molecules across the newly formed complex surfaces. Presumably, the polar ligand surface promotes the formation of more transient interactions with solvating water molecules that cannot be observed with crystallography. A comparable hydrophobic moiety of equivalent size seems to establish a more defined and highly directional network of capping water molecules that takes beneficial impact on the thermodynamic signature of the binding process.

Most importantly, all of the polar inhibitors investigated in this study display a considerably lower affinity to TLN compared to the more hydrophobic phosphonopeptide analogues lacking the terminal ammonium group. This is probably owed to the costly partial desolvation of the charged ammonium group in **1**–**8**. Even when the polar group is able to engage in hydrogen bond interactions with the protein or succeeds to recruit several surrounding water molecules, a pronounced disadvantage in the free energy of binding is recorded for the more hydrophilic inhibitor analogues. This is even true in the case of **5**, where the polar terminal group sticks out into the solvent and is not involved in any direct interactions to the protein or water molecules of the first solvation layer.

A direct comparison of the binding characteristics of the polar inhibitors **3**–**4** and their hydrophobic analogues **9**–**10** is given in Figure 4.3A. The overall binding mode of the inhibitor scaffold is virtually conserved for all examples. The deviating orientations of the P_2' side chains, however, match with the changes of the chemical properties of the respective inhibitors. Whereas **9** and **10** occupy the hydrophobic cavity below Asn111, the terminal ammonium groups in **3** and **4** are oriented in a way that the polar head group is integrated into a network of surrounding water molecules. Even though the P_2' side chains exhibit apolar properties, their hydration in TLN-**9** and TLN-**10** results in sophisticated solvation network patterns. In both cases, more water molecules wrap around the exposed ligand portion than in TLN-**3** and TLN-**4** respectively. Similar conclusions can be drawn for the comparison of **7** and **8** to their respective analogues **11** and **12** (Figure 4.3B).

The apolar derivatives show a gain of affinity by a factor of 20–180 in K_d when compared to their polar analogs. The observation that this effect seems to originate from a strong enthalpic advantage is in line with the hypothesis that the rather weak affinity of **3**–**4** and **7**–**8** can be attributed to an unfavorable desolvation penalty of the polar head group. Placing the polar

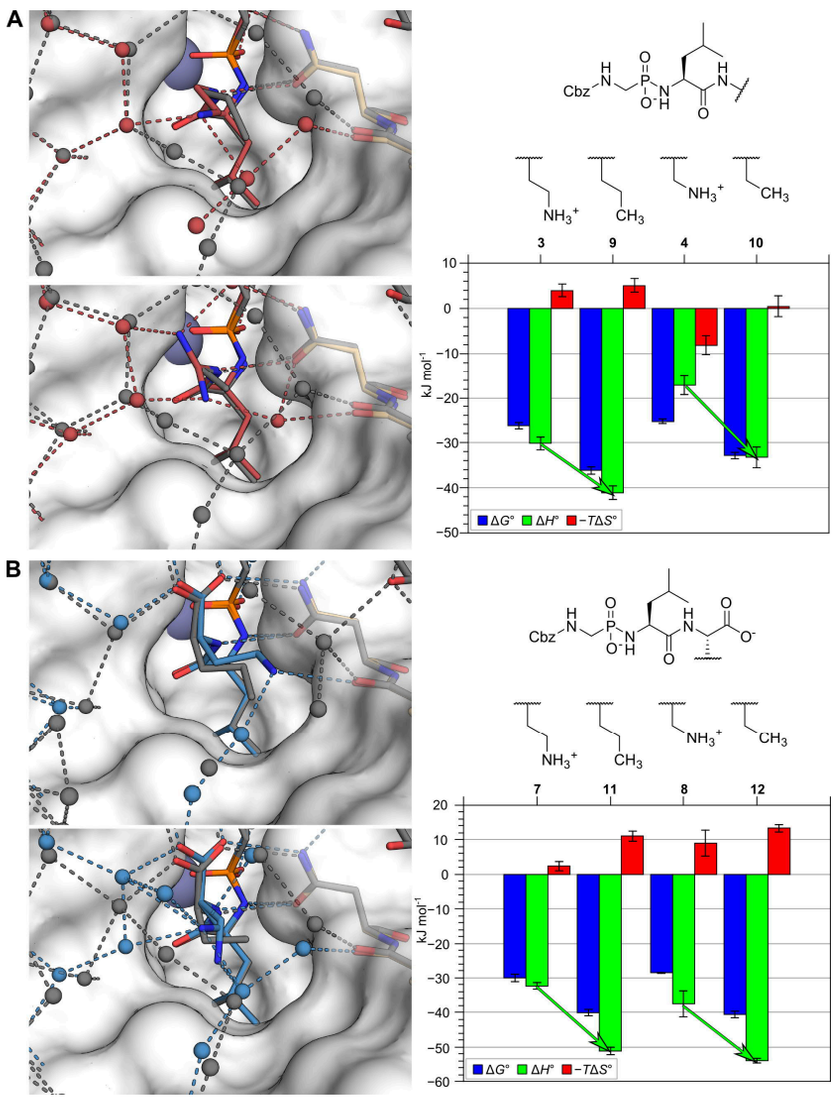


Figure 4.3. Comparison of binding characteristics of polar inhibitors **3**, **4**, **7**, **8** and aliphatic compounds **9**,**10**,**11**,**12**. (A) Superposition of the crystal structures TLN-**3**/TLN-**9** and TLN-**4**/TLN-**10** and thermodynamic binding profiles of the respective inhibitors. (B) Superposition of the crystal structures TLN-**7**/TLN-**11** and TLN-**8**/TLN-**12** and thermodynamic binding profiles of the respective inhibitors. Carbon atoms, water molecules and hydrogen bonds of TLN-**3** and TLN-**4** are colored in red (protein residues in beige), carbon atoms, water molecules and hydrogen bonds of TLN-**7** and TLN-**8** are colored in blue (protein residues in beige), carbon atoms, water molecules and hydrogen bonds of TLN-**9**, TLN-**10**, TLN-**11**, and TLN-**12** are shown in gray. Heteroatoms are color coded according to atom type and the solvent excluded protein surface is depicted in white.

group more remotely by using a longer chain as in **1**–**2** and **5**–**6**, makes the ligand slightly more potent, the enthalpic cost of desolvation, however, is still to be afforded. In all cases, the purely hydrophobic group oriented toward the solvent reveals the more potent binding. This is

emphasized in cases, where the substituent introduces a sophisticated and well-packed network of water molecules.

Remarkably, no entropic advantage is recorded for the ligands bearing a strictly hydrophobic alkyl chain. This observation further underlines the fact that simplified models of the entropy driven "hydrophobic effect", which are ultimately based on studies of simple hydrophobic molecules in aqueous solution [6], are not sufficient when complex topologies of protein–ligand interfaces are considered [14].

4.5 Conclusion

The present study provides a number of important insights. The S_2' pocket of TLN is a flat, mainly hydrophobic pocket that opens to the solvent environment. Intuitively, such pockets are not considered as important in the optimization of ligand binding. At best, such pockets, where ligand portions orient away from the protein binding site, are seen as ideal to attach polar solubilizing groups to improve pharmacokinetic properties within a ligand class. In a previous study, we could show that attachment of apolar P_2' substituents, which orient into the interface between protein and surrounding water bulk phase, can enhance ligand affinity up to 50 fold. The affinity gain is maximized, if the network of water molecules of the first solvation shell adopts an ideal, energetically favored architecture. The present example underlines, particularly by facing matching pairs of polar and apolar P_2' substituents, that the attachment of a terminal ammonium group reduces the affinity of the considered ligands by a factor of up to 180 in K_d

Even when the polar moiety remains partly solvated or engages in strong interactions to the protein, an enthalpic penalty, compared to its apolar analogue, is observed. Surprisingly, the positioning of the charged group close to the solvent interface of the protein–ligand complex did not induce a pronounced ordering leading to a strong organization of the surrounding hydration shell. The interactions of solvent molecules with the charged ammonium group seem to be of a less structured and more transient nature, rendering them difficult to observe by X-ray crystallography. In terms of ligand potency, the partial desolvation of a hydrophobic group in conjunction with an optimized structuring of a distinct water network in its hydration shell, was found to be more beneficial than the introduction of a terminal polar substituent, for which the costs of partial desolvation have to be accounted for. Within the studied series, the enthalpic costs that have to be paid for this effect were largely independent of the chain length to which the charged moiety was attached. We believe that it is reasonable to assume that a similar energetic inventory will be observed for other systems, where ligand portions bind next to a solvent interface. The study of such residual hydration effects next to the first solvation layer will help to improve our understanding of the basic principles of ternary protein–ligand–

solvent interactions and may prove useful for the developement of more effective drug molecules.

4.6 Experimental Section

4.6.1 Inhibitor Synthesis

The synthesis of inhibitors **1**–**12** is detailed elsewhere [15,93,183].

4.6.2 ITC Data Collection

Thermodynamic profiles of all inhibitors were recorded on a Microcal ITC200 (GE Healthcare) or Microcal ITC200 (Malvern) device. TLN was acquired from Calbiochem (EMD Biosciences) as a freeze dried powder. Protein samples for ITC measurements were prepared by dissolving the freeze dried powder in the respective buffer (20 mM HEPES, PIPES or sodium cacodylate, 2 mM $CaCl_2$, 500 mM NaSCN, pH7.5) directly before the experiment. A protein concentration of 250 µM was used for the experiments with **1**–**8**. The high protein concentration was necessary to achieve a c-value > 1 ($c = n[P]_0 K_a$) for low affinity compounds. Inhibitor solutions were prepared from the freeze dried solids. A titrand concentration between 3 mM and 3.5 mM with 10–15 injections of 2.4–3.5 µL was used for the titrations. For synthetic reasons, ligand **8** could only be obtained in a diastereomeric mixture with its P_2' epimer. Since the lack of saturation and the binding stoichiometry (n) in initial titrations indicated that the (*S,R*)-diastereomer did not bind to TLN, the concentration was adjusted to 7 mM for this ligand, assuming that the interaction is predominantly determined by the (*S,S*)-diastereomer. The purity of all ligands was determined by qNMR and the concentrations were corrected accordingly. Thermodynamic binding profiles of **9**–**12** in HEPES buffer were first published in earlier contributions [15,93]. Due to varying experimental conditions (buffer composition, protein batch, experimentator), the experiments were repeated for this study and a buffer correction was performed. For these compounds, a protein concentration of 48 µM and a ligand concentration of 1 mM with 10–15 injections of 1.6–3.5 µL was used. The raw thermograms were analyzed with *NITPIC* [167] version 1.1.2 and *SEDPHAT* [184] version 12.1b. All titration were performed in duplicate or triplicate.

4.6.3 Crystal Preparation and Soaking

To an 8 mM solution of lyophilized TLN powder (Calbiochem) in 50 µL DMSO, 50 µL of a solution containing 3.8 M CsCl and 100 mM Tris-HCl (pH 7.5) was added. After

centrifugation (3 min, 16,000 g), 1 µL of the clear solution was transferred into the wells of a 24 well sitting drop crystallization plate. The reservoir wells were filled with 1 mL of demineralized water. After sealing the crystallization plates, the crystals finished growing after 5 days at 18 °C. Soaking of the ligands was performed at a concentration of 1 mM for 24 h in a buffer composed of 100 mM Tris-HCl (pH 7.5), 2 mM $CaCl_2$, 5% DMSO. Subsequently, the crystals were flash-frozen in liquid nitrogen in a buffer composed of 100 mM Tris-HCl (pH 7.5), 10 mM $CaCl_2$, 5% DMSO, 20% MPD and 1 mM of the respective ligand.

4.6.4 Data Collection, Processing, Structure Determination and Refinement

Data collection for the TLN-**1** to TLN-**8** crystal structures was performed at BESSY II (Helmholtz-Zentrum Berlin) MX-beamline 14.1 at a wavelength of 0.91841 Å. This facility is equipped with a Dectris Pilatus 6M pixel detector. For indexing, integration and scaling of the datasets, *XDS* [185] was used. Molecular replacement using *Phaser* (version 2.5.0) [159] was applied for phasing with PDB 8TLN as a search model. Cartesian simulated annealing was performed (default settings) as a first refinement step. Refinement of the initial model (xyz coordinates, individual *B* factors, occupancies) was performed with *Phenix.refine* (version 1.10.1-2155) [162] and *Coot* (version 0.7) [186] was used for model building into σ-weighted maps ($2F_o–F_c$ and $F_o–F_c$). Alternating cycles of these processes were performed until *R* values reached convergence. For the calculation of R_{free}, a randomly chosen subset of 5% of the reflections was excluded from the refinement. Hydrogen atoms were included to the model with *Phenix.refine*. With the exception of hydrogen atoms, an anisotropic refinement strategy of *B* factors for all model atoms was applied. An occupancy cutoff of 20% was used for alternative conformations of protein residues. Ligand molecules were created with *MOE* [151] and restraints were calculated with *Phenix.elbow* [163]. Data collection and refinement statistics are given in Table 4.1 and Table 4.2.

4.7 Accession Codes

Atomic coordinates and experimental details for all new crystal structures will be released in the PDB upon publication (TLN-**1**: 5N2Z; TLN-**2**: 5N2T; TLN-**3**: 5N34; TLN-**4**: 5N3V; TLN-**5**: 5N31; TLN-**6**: 5N2X; TLN-**7**: 5MNR; TLN-**8**: 5N3Y). TLN-**9** (4MXJ), TLN-**10** (3T8F), TLN-**11** (3T8C), and TLN-**12** (3T87) were reported earlier [15,93].

4.8 Acknowledgement

The authors want to thank the beamline support team at BESSY II (Helmholtz-Zentrum Berlin, Germany) for helpful advice during data collection. The authors acknowledge the receipt of a travel grant from the Helmholtz-Zentrum Berlin. This work was funded by the European Research Council (ERC) of the European Union (grant 268145-DrugProfilBind).

4.9 Supporting Information

4.9.1 Results From Individual ITC Titrations

Table S4.1. Thermodynamic data from individual ITC measurements.

Measurement	*n*	K_d (µM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
Cacodylate-**1** a	1.12	5.02	-30.2	-33.3	3.0
Cacodylate-**1** b	1.09	5.15	-30.2	-31.8	1.6
Cacodylate-**1** c	1.12	4.54	-30.5	-32.1	1.6
	1.11±0.02	**4.90±0.32**	**-30.3±0.2**	**-32.4±0.8**	**2.1±0.8**
HEPES-**1** a	0.98	10.93	-28.3	-14.1	-14.2
HEPES-**1** b	1.05	8.33	-29.0	-14.1	-14.9
HEPES-**1** c	1.09	9.03	-28.8	-14.0	-14.8
	1.04±0.06	**9.43±1.35**	**-28.7±0.3**	**-14.1±0.1**	**-14.6±0.4**
PIPES-**1** a	1.09	5.92	-29.8	-21.1	-8.8
PIPES-**1** b	1.00	6.74	-29.5	-18.8	-10.7
PIPES-**1** c	1.03	6.12	-29.8	-20.4	-9.3
	1.04±0.05	**6.26±0.43**	**-29.7±0.2**	**-20.1±1.2**	**-9.6±1.0**
Cacodylate-**2** a	1.00	8.64	-28.9	-32.0	3.1
Cacodylate-**2** b	0.95	8.58	-28.9	-32.4	3.4
Cacodylate-**2** c	0.97	7.99	-29.1	-32.1	3.0
	0.97±0.03	**8.40±0.35**	**-29.0±0.1**	**-32.2±0.2**	**3.2±0.2**
HEPES-**2** a	0.89	12.87	-27.9	-15.1	-12.8
HEPES-**2** b	0.88	12.28	-28.0	-15.3	-12.7
HEPES-**2** c	0.84	15.01	-27.5	-15.3	-12.2
	0.87±0.03	**13.39±1.44**	**-27.8±0.3**	**-15.2±0.1**	**-12.6±0.3**

Table S4.1. Continued.

Measurement	*n*	K_d (μM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
PIPES-**2** a	0.91	10.27	-28.5	-24.9	-3.5
PIPES-**2** b	0.73	11.33	-28.2	-27.3	-1.0
PIPES-**2** c	0.83	11.59	-28.2	-26.5	-1.6
	0.82±0.09	**11.06±0.70**	**-28.3±0.2**	**-26.2±1.2**	**-2.0±1.3**
Cacodylate-**3** a	0.80	24.50	-26.3	-32.2	5.9
Cacodylate-**3** b	0.73	22.03	-26.6	-33.0	6.4
	0.77±0.05	**23.27±1.74**	**-26.5±0.2**	**-32.6±0.6**	**6.1±0.4**
HEPES-**3** a	0.69	30.67	-25.8	-13.2	-12.6
HEPES-**3** b	0.65	23.47	-26.4	-14.5	-11.9
HEPES-**3** c	0.57	32.03	-25.7	-8.6	-17.1
	0.64±0.06	**28.72±4.60**	**-25.9±0.4**	**-12.1±3.1**	**-13.9±2.8**
PIPES-**3** a	0.78	28.54	-25.7	-8.6	-17.1
PIPES-**3** b	0.79	32.82	-25.9	-22.2	-3.8
PIPES-**3** c	0.78	19.24	-25.6	-23.0	-2.6
	0.78±0.01	**26.87±6.94**	**-26.1±0.7**	**-22.0±1.1**	**-4.2±1.8**
Cacodylate-**4** a	0.74	41.41	-25.0	-18.6	-6.4
Cacodylate-**4** b	0.78	46.40	-24.7	-18.1	-6.6
Cacodylate-**4** c	0.84	47.92	-24.7	-18.3	-6.4
	0.79±0.05	**45.24±3.41**	**-24.8±0.2**	**-18.3±0.3**	**-6.5±0.2**
HEPES-**4** a	0.44	22.22	-26.6	-4.2	-22.3
HEPES-**4** b	0.44	18.43	-27.0	-3.9	-23.2
HEPES-**4** c	0.40	24.29	-26.3	-4.6	-21.7
	0.43±0.02	**21.65±2.97**	**-26.6±0.3**	**-4.2±0.4**	**-22.4±0.7**
PIPES-**4** a	0.64	54.85	-24.3	-12.8	-11.5
PIPES-**4** b	0.50	47.33	-24.7	-13.3	-11.3
PIPES-**4** c	0.61	47.35	-24.7	-12.4	-12.3
	0.58±0.07	**49.84±4.34**	**-24.6±0.2**	**-12.9±0.5**	**-11.7±0.5**
Cacodylate-**5** a	0.97	1.44	-33.3	-40.8	7.5
Cacodylate-**5** b	0.84	1.86	-32.7	-40.5	7.8
Cacodylate-**5** c	0.92	1.53	-33.2	-41.1	7.9
	0.91±0.07	**1.61±0.22**	**-33.1±0.3**	**-40.8±0.3**	**7.7±0.2**

Table S4.1. Continued.

Measurement	*n*	K_d (μM)	Δ*G*° (kJ mol^{-1})	Δ*H*° (kJ mol^{-1})	−*T*Δ*S*° (kJ mol^{-1})
HEPES-**5** a	0.88	2.44	-32.0	-21.2	-10.8
HEPES-**5** b	0.87	2.55	-31.9	-22.0	-9.9
	0.88±0.01	**2.49±0.08**	**-32.0±0.1**	**-21.6±0.6**	**-10.4±0.7**
PIPES-**5** a	0.96	1.77	-32.8	-28.4	-4.5
PIPES-**5** b	0.82	1.29	-33.6	-29.7	-3.9
PIPES-**5** c	0.90	1.54	-33.2	-29.8	-3.3
	0.89±0.07	**1.53±0.24**	**-33.2±0.4**	**-29.3±0.8**	**-3.9±0.6**
Cacodylate-**6** a	1.02	0.68	-35.2	-37.0	1.8
Cacodylate-**6** b	0.99	0.76	-34.9	-40.0	5.1
Cacodylate-**6** c	1.00	1.18	-33.8	-40.8	6.9
	1.00±0.01	**0.87±0.27**	**-34.7±0.7**	**-39.3±2.0**	**4.6±2.6**
HEPES-**6** a	1.00	2.81	-31.7	-22.9	-8.8
HEPES-**6** b	1.03	1.96	-32.6	-22.2	-10.4
HEPES-**6** c	1.00	1.05	-34.1	-21.1	-13.0
	1.01±0.02	**1.94±0.88**	**-32.8±1.2**	**-22.1±0.9**	**-10.7±2.1**
PIPES-**6** a	0.91	0.71	-35.1	-30.2	-4.9
PIPES-**6** b	0.98	1.57	-33.1	-31.4	-1.7
PIPES-**6** c	0.97	1.59	-33.1	-33.1	0.0
	0.95±0.03	**1.29±0.50**	**-33.8±1.1**	**-31.6±1.5**	**-2.2±2.5**
v	0.95	4.33	–30.6	–36.2	5.5
Cacodylate-**7** b	0.93	2.55	–31.9	–32.9	1.0
	0.94±0.02	**3.44±1.26**	**–31.3±0.9**	**–34.5±2.3**	**3.3±3.2**
HEPES-**7** a	0.87	11.42	–28.2	–17.4	–10.9
HEPES-**7** b	0.88	9.47	–28.7	–17.6	–11.1
HEPES-**7** c	0.79	7.69	–29.2	–17.1	–12.1
	0.85±0.05	**9.53±1.87**	**–28.7±0.5**	**–17.4±0.3**	**–11.3±0.7**
PIPES-**7** a	0.80	5.45	–30.0	–25.6	–4.4
PIPES-**7** b	0.84	4.50	–30.5	–25.8	–4.7
PIPES-**7** c	0.82	5.78	–29.9	–24.7	–5.2
	0.82±0.02	**5.24±0.67**	**–30.2±0.3**	**–25.4±0.6**	**–4.8±0.4**

Table S4.1. Continued.

Measurement	*n*	K_d (μM)	Δ*G*° (kJ mol^{-1})	Δ*H*° (kJ mol^{-1})	−*T*Δ*S*° (kJ mol^{-1})
Cacodylate-**8** a	0.85	9.53	−28.7	−38.4	9.7
Cacodylate-**8** b	0.90	9.42	−28.7	−38.4	9.7
Cacodylate-**8** c	0.78	9.00	−28.8	−39.5	10.7
	0.84±0.02	**9.32±0.28**	**−28.7±0.1**	**−38.7±0.6**	**10.0±0.5**
HEPES-**8** a	0.82	9.83	−28.6	−18.3	−10.3
HEPES-**8** b	0.72	11.40	−28.2	−20.7	−7.5
HEPES-**8** c	0.73	11.49	−28.2	−19.8	−8.4
	0.76±0.06	**10.91±0.93**	**−28.3±0.2**	**−19.6±1.2**	**−8.7±1.4**
PIPES-**8** a	0.71	9.97	−28.5	−31.9	3.3
PIPES-**8** b	0.75	10.32	−28.5	−33.4	5.0
	0.73±0.03	**10.15±0.25**	**−28.5±0.1**	**−32.7±1.1**	**4.2±1.1**
Cacodylate-**9** a	1.02	0.34	−36.9	−43.7	6.8
Cacodylate-**9** b	1.08	0.30	−37.2	−44.1	6.9
	1.05±0.05	**0.32±0.03**	**−37.1±0.2**	**−43.9±0.3**	**6.9±0.1**
HEPES-**9** a	0.98	0.59	−35.6	−19.7	−15.9
HEPES-**9** b	0.97	0.67	−35.3	−20.4	−14.8
	0.98±0.01	**0.63±0.05**	**−35.4±0.2**	**−20.1±0.5**	**−15.4±0.8**
PIPES-**9** a	0.99	0.46	−36.2	−31.6	−4.5
PIPES-**9** b	1.02	0.53	−35.8	−32.2	−4.4
	1.01±0.02	**0.50±0.05**	**−36.0±0.3**	**−31.5±0.2**	**−4.5±0.1**
Cacodylate-**10** a	1.02	1.36	−33.5	−35.6	2.1
Cacodylate-**10** b	0.99	1.19	−33.8	−35.1	1.3
	1.01±0.02	**1.27±0.12**	**−33.7±0.2**	**−35.3±0.4**	**1.7±0.6**
HEPES-**10** a	0.97	2.30	−32.2	−14.1	−18.1
HEPES-**10** b	0.94	2.35	−32.1	−14.2	−17.9
	0.95±0.02	**2.33±0.04**	**−32.2±0.1**	**−14.2±0.1**	**−18.0±0.1**
PIPES-**10** a	1.08	2.03	−32.5	−26.3	−6.2
PIPES-**10** b	1.03	1.86	−32.7	−25.5	−7.2
	1.06±0.04	**1.94±0.11**	**−32.6±0.2**	**−25.9±0.6**	**−6.7±0.7**
Cacodylate-**11** a	1.12	0.07	−40.8	−53.7	12.9
Cacodylate-**11** b	1.18	0.06	−41.4	−55.5	14.1
	1.15±0.05	**0.06±0.01**	**−41.1±0.4**	**−54.6±1.2**	**13.5±0.9**

Table S4.1. Continued.

Measurement	*n*	K_d (μM)	Δ*G*° (kJ mol^{-1})	Δ*H*° (kJ mol^{-1})	−*T*Δ*S*° (kJ mol^{-1})
HEPES-**11** a	1.07	0.13	–39.3	–27.5	11.8
HEPES-**11** b	1.09	0.17	–38.6	–28.5	10.2
	1.08±0.01	**0.15±0.03**	**–39.0±0.5**	**–28.0±0.7**	**11.0±1.2**
PIPES-**11** a	1.01	0.09	–40.3	–40.0	–0.3
PIPES-**11** b	1.11	0.09	–40.3	–40.0	–0.2
	1.06±0.07	**0.09±0.00**	**–40.3±0.0**	**–40.0±0.0**	**–0.3±0.1**
Cacodylate-**12** a	1.24	0.06	–41.1	–58.5	17,4
Cacodylate-**12** b	1.19	0.05	–41.6	–55.7	14.2
	1.21±0.03	**0.06±0.01**	**–41.3±0.3**	**–57.1±1.9**	**15.8±2.3**
HEPES-**12** a	1.18	0.11	–39.7	–33.3	–6.4
HEPES-**12** b	1.22	0.13	–39.2	–34.3	–5.0
	1.20±0.03	**0.12±0.02**	**–39.4±0.3**	**–33.8±0.7**	**–5.7±1.0**
PIPES-**12** a	1.16	0.06	–41.0	–42.7	–1.7
PIPES-**12** b	1.20	0.07	–40.8	–45.1	–4.3
	1.18±0.03	**0.07±0.00**	**–40.9±0.2**	**–43.9±1.7**	**–3.0±1.9**

4.9.2 Representative Thermograms and Binding Isotherms

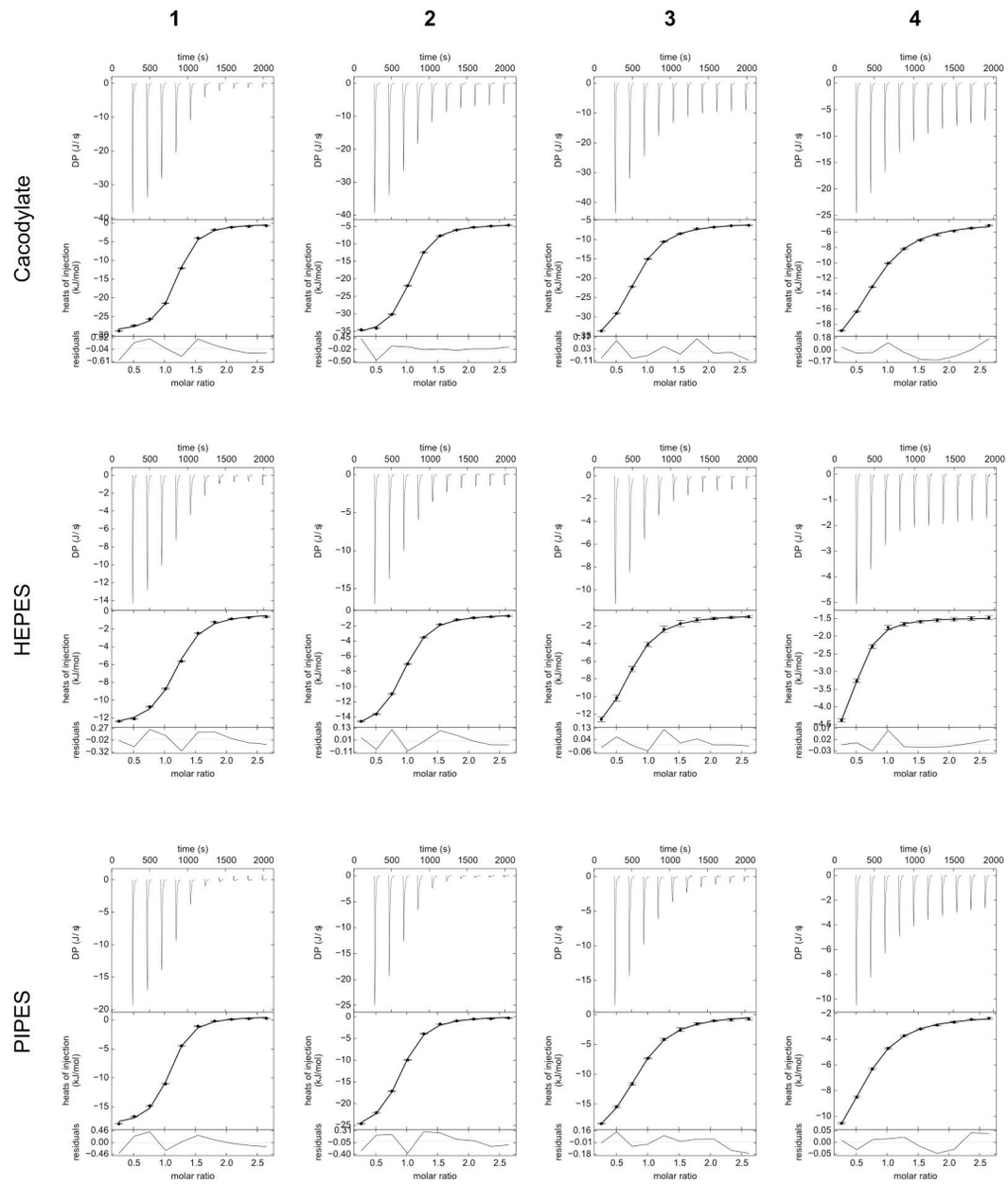


Figure S4.1. Representative thermograms and extracted binding isotherms from ITC measurements of **1**–**4**..

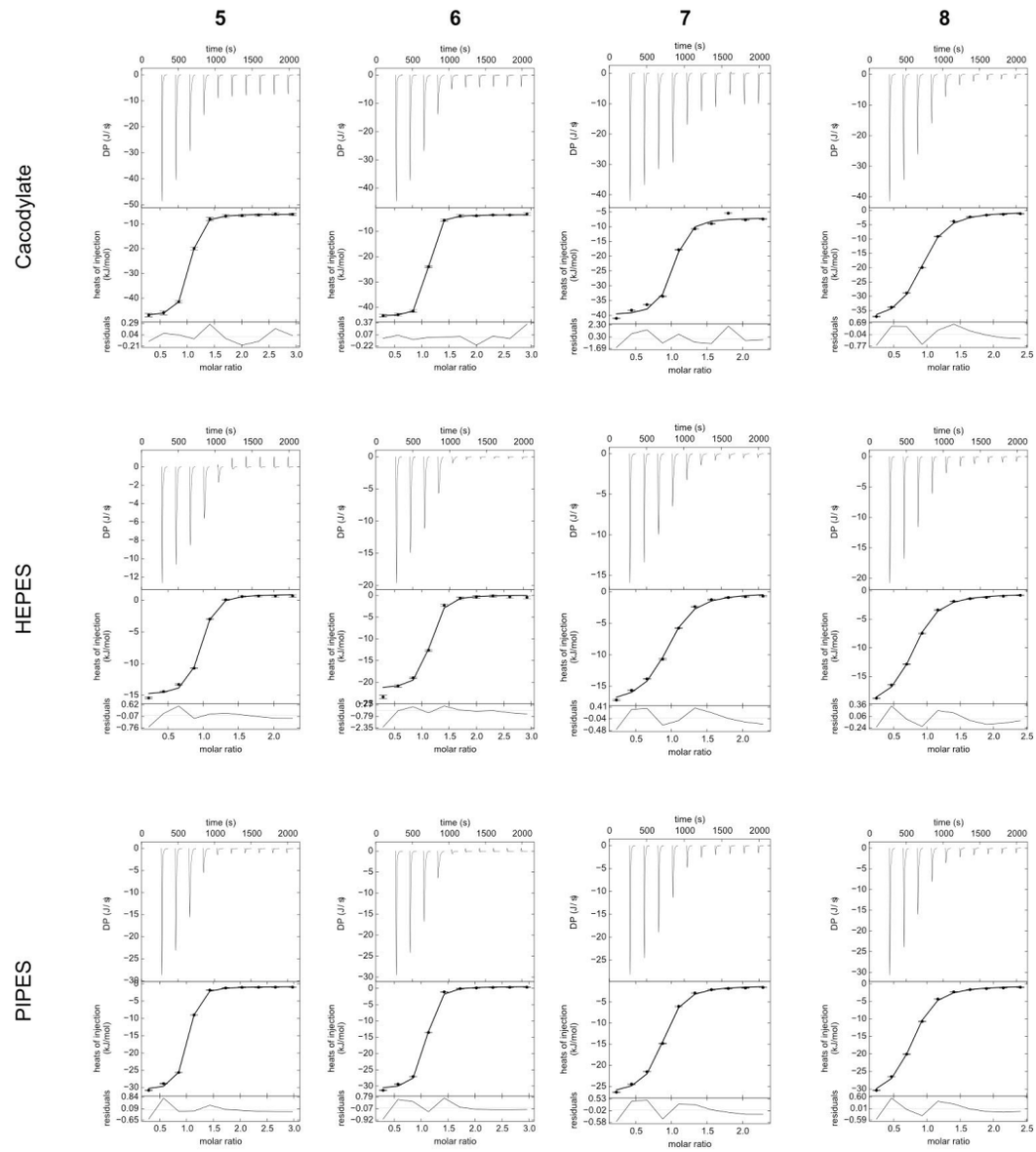


Figure S4.2. Representative thermograms and extracted binding isotherms from ITC measurements of **5**–**8**..

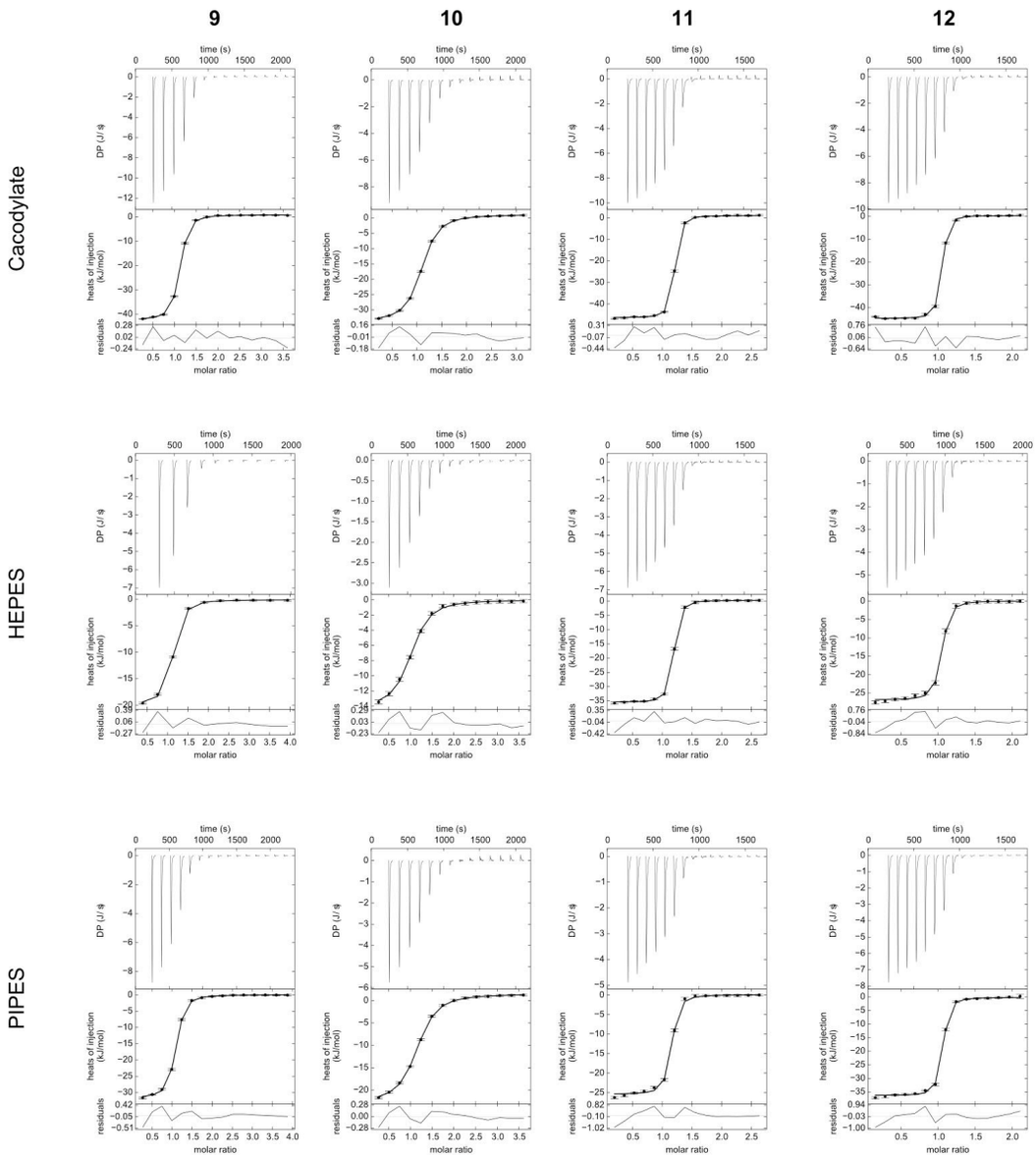


Figure S4.3. Representative thermograms and extracted binding isotherms from ITC measurements of **9**–**12**..

4.9.3 **Buffer correction of the thermodynamic binding profiles**

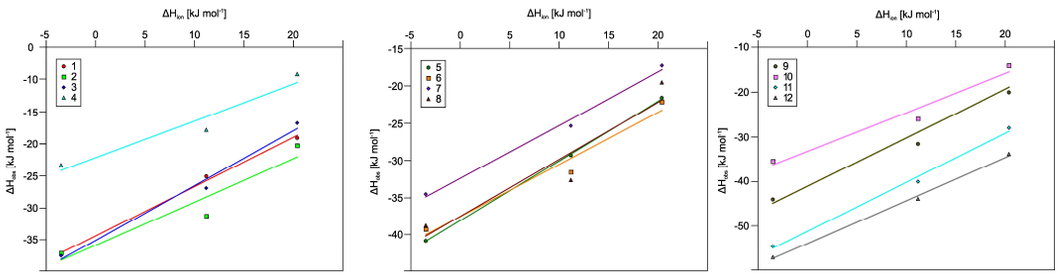


Figure S4.4. Plot of the observed binding enthalpy in three buffers against ionization enthalpies (taken from [297]).

Table S4.2. Buffer corrected thermodynamic data.

Inhibitor	Δn[a]	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
1	0.77±0.05	−29.6±0.4	−29.4±0.7	−0.15±0.8
2	0.68±0.21	−28.4±0.3	−30.9±2.9	2.6±2.9
3	0.86±0.10	−26.2±0.7	−30.1±1.4	3.9±1.6
4	0.57±0.15	−25.3±0.5	−17.2±2.1	−8.2±2.2
5	0.80±0.01	−32.8±0.5	−38.1±0.2	5.3±0.5
6	0.70±0.14	−33.7±1.8	−37.5±1.9	3.8±2.6
7	0.71±0.07	−30.0±1.1	−32.4±0.9	2.4±1.4
8	0.77±0.27	−28.5±0.2	−37.5±3.7	9.0±3.7
9	0.98±0.11	−36.1±0.8	−41.1±1.5	5.0±1.7
10	0.86±0.17	−32.8±0.7	−33.2±2.3	0.4±2.4
11	1.10±0.08	−40.1±1.0	−51.2±1.1	11.1±1.5
12	0.97±0.05	−40.6±0.9	−54.0±0.7	13.4±1.1

[a] Number of protons transferred during ligand binding. Inferred from the slope of the regression line of the plot of ΔH_{obs} against ΔH_{ion}.

4.9.4 Additional Crystallographic Observations

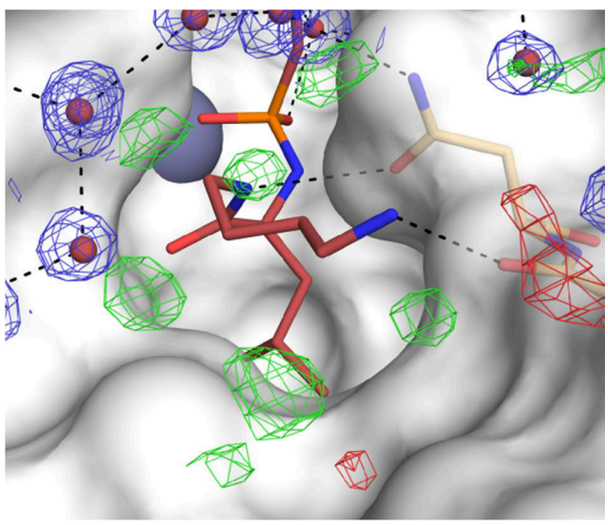


Figure S4.5. Residual electron density in the crystal structure of **1**. The 2mF_o–DF_c electron density for the water molecules that were included in the crystallographic models is shown as a blue mesh (1.0σ). The residual electron density (mF_o–DF_c density map, 3.0σ) is shown as green/red meshes for positive/negative peaks respectively.

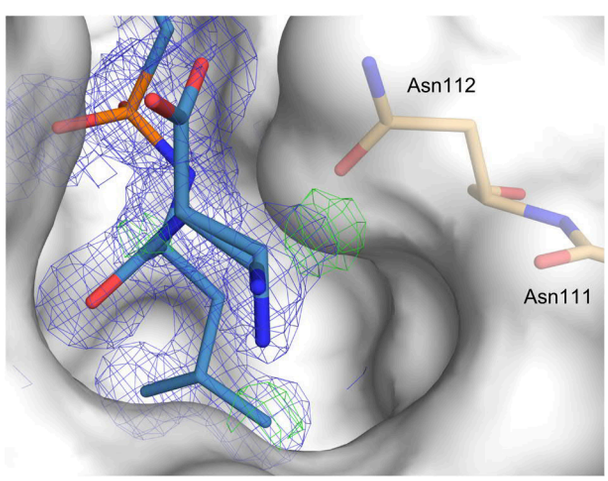


Figure S4.6. Indication of a third conformation of the ligand side chain in the crystal structure of **8**. The 2mF_o–DF_c (1.0σ) and the mF_o–DF_c (3.0σ) electron density maps of the ligand are shown as blue and green meshes, respectively. Water molecules and hydrogen bond interactions are omitted for reasons of clarity.

Chapter 5

How Nothing Boosts Affinity: Hydrophobic Ligand Binding to the Virtually Vacated S_1' Pocket of Thermolysin

Stefan G. Krimmer,[1,*] Jonathan Cramer,[1,*] Johannes Schiebel,[1]
Andreas Heine,[1] Gerhard Klebe[1]

*These authors contributed equally

[1]Department of Pharmaceutical Chemistry, University of Marburg,
Marbacher Weg 6, 35032 Marburg, Germany

Manuscript in preparation for submission

5.1 Introduction

The deep, hydrophobic S_1' pocket of thermolysin (TLN), which is large enough to accommodate a P_1' substituent up to a benzyl side chain, has previously been reported by us to be poorly solvated and to contain only disordered water molecules, whose displacement would give rise to a strong enthalpic benefit [33]. In the current contribution, we try to collect more evidence for this poor solvation pattern, maximally composed of disordered water molecules in the S_1' pocket of this metalloprotease, and investigate the origin of the reported enthalpy gain. In this context, we analyze the protein–ligand complex formation of a series of six phosphonamidate-type TLN inhibitors (Figure 5.1A) using X-ray crystallography and isothermal titration calorimetry (ITC). Going from a hydrogen atom to a *iso*-butyl P_1' group (Figure 5.1B), an amazingly strong affinity increase of 26.3 kJ mol^{-1} is observed. For the explanation of the molecular mechanism behind this considerable affinity increase, exact knowledge about the solvation state within the S_1' cavity is required. For TLN-**1**, we determined an experimentally phased electron density map via a zinc multi-wavelength anomalous diffraction (MAD) experiment. The experimentally phased electron density map provides the advantage that no model bias is introduced by phasing of the structure compared to less elaborate strategies such as molecular replacement. Additionally, we attempt to transform the sigma-scaled electron density map to an absolute electron number scale ($e^-/Å^3$) to enable the determination of the exact total electron content within the S_1' cavity. This strategy allows the detection of electron density traces, which can indicate highly mobile, crystallographically disordered water molecules [187]. To characterize the volume and polarity of the S_1' cavity of TLN-**1** experimentally, we used the noble gases xenon and krypton as probes to fill the pocket. Considering the complexes of **1** to **4**, the residual unoccupied volume of the S_1' cavity is gradually reduced, thus decreasing the putative residence volume of a potentially bound mobile water molecule. Inhibitors **5** and **6** even comprise polar functional groups that increase the polarity within the cavity and provide a potential hydrogen-bonding anchor to fix a remaining S_1' water occupant. Both strategies, reducing the residence volume and increasing the polarity within the cavity, should diminish the mobility of a potentially present residual water molecule and increase the concentration of its electrons within the cavity, thus facilitating its detection in a crystal structure.

5.2 Results

5.2.1 Isothermal titration calorimetry

ITC measurements of **2**–**4** were performed by direct titrations (Figure S5.1A–C) and for **1**, **5** and **6** by displacement titrations (Figure S5.1D–F). The thermodynamic profiles of **2**–**4** were

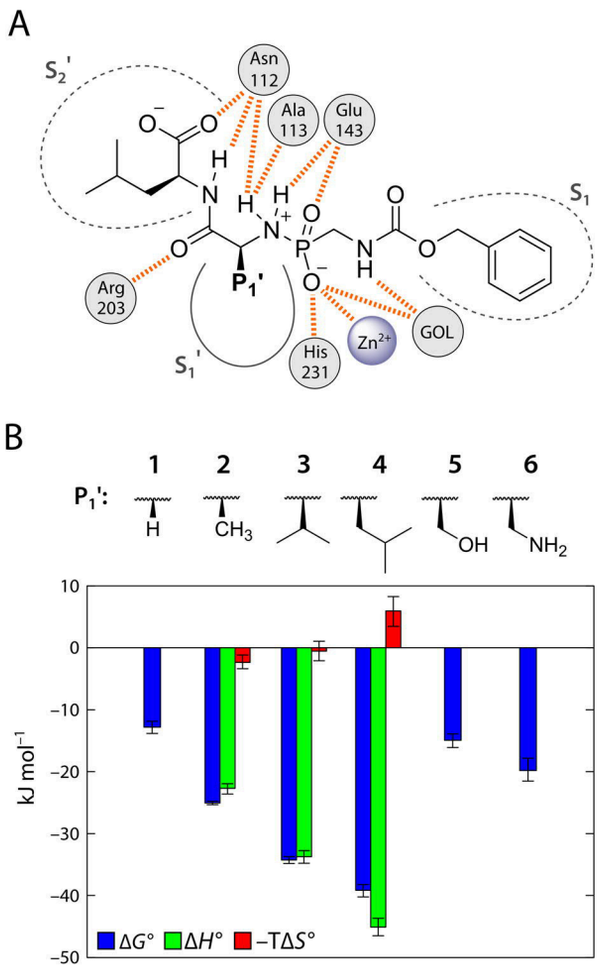


Figure 5.1. Congeneric series of phosphonamidate TLN ligands substituted with different P_1' groups. **(A)** Schematic binding mode of the parent ligand scaffold in complex with TLN. Protein residues and the zinc ion forming hydrogen bonds (orange dashed lines) with the parent scaffold are indicated (GOL = glycerol molecule from the cryo buffer bound to TLN). The $S_1$and S_2' and pockets of TLN are wide-open and well-accessible to water molecules, whereas the S_1' pocket is deep and apolar. **(B)** P_1' substituents of **1**–**6** and their thermodynamic binding profiles as determined by ITC. The thermodynamic profiles shown for **2**, **3** and **4** were corrected for the heat of buffer ionization upon complex formation. Error bars represent the 95.4% confidence interval. A reliable determination of the buffer-corrected enthalpy/entropy partitioning of **1**, **5** and **6** (displacement titrations) is impossible and thus in these cases only the standard Gibbs free energy is discussed. Data values are listed in Table S5.1.

corrected for proton transfer reactions upon complex formation via measurement in several buffers showing different ionization enthalpies (Figure S5.2) [123]. Such corrections were not performed for the displacement titrations of **1**, **5** and **6**, because this would have resulted in large experimental uncertainties, rendering the thermodynamic parameters unreliable even for mutual comparison.

The thermodynamic profiles of **1**–**6** are displayed in Figure 5.1B. The affinity strongly increases with the growing hydrophobic size of the P_1' group (**1**→**2**→**3**→**4**) over 4.5 orders of magnitude (expressed in terms of the binding constants as listed in Table S5.1). In particular, the addition of a single methyl group to **1** yielding **2** ($\Delta\Delta G°_{1\to2}$= −12.3 kJ mol^{-1}) and the addition of two methyl groups to **2** yielding **3** ($\Delta\Delta G°_{2\to3}$ = −9.2 kJ mol^{-1}) result in substantial affinity boosts. With respect to the partitioning in $\Delta H°$ and $-T\Delta S°$, the affinity increase from **2**→**3**→**4** is the result of a sharp increase in $\Delta H°$, which is only partially compensated by a decrease in $-T\Delta S°$. The affinities of **5** and **6** fall between those of **1** and **2**. The amino-substituted **6** clearly shows a higher affinity than the hydroxyl-substituted **5**.

5.2.2 Crystal structure analysis

In addition to the already published crystal structures of TLN-**2**, TLN-**3** and TLN-**4** [33], we succeeded in crystallizing TLN-**1**, TLN-**5**, and TLN-**6** (Table 5.1). Furthermore, we also obtained TLN-**1** in complex with xenon (TLN-**1**-Xe) and krypton (TLN-**1**-Kr). Except for the native structure of TLN-**1** that was experimentally phased, all other crystal structures were phased by the standard molecular replacement technique.

5.2.3 Shape, polarity and solvent content analysis within the S_1' cavity and within three internal reference cavities of TLN-1

The structure of TLN-**1** was experimentally phased using a zinc MAD dataset (Table 5.2). The experimental phasing of TLN-**1** without any further least-squares refinement steps resulted in a very clear, high-quality electron density (Supporting Information, Figure S5.3) with an overall figure of merit (FOM) of 0.870. We decided to apply density modification techniques (solvent flattening and histogram matching) on the experimental phases since this resulted in a significantly improved quality of the electron density map (overall FOM increased from 0.763 to 0.870). Since the analyzed cavities are completely buried within the core of TLN and exhibit a narrow shape, the electron density within the cavities is not affected by the applied phase improvement techniques. The experimentally phased density map was put on an absolute electron number scale (see Chapter 5.5.8 for further details), and the total electron content within the S_1' cavity and three internal reference cavities of different solvation state was calculated. The protein model coordinates displayed in Figure 5.2A–E are taken from the experimentally phased, fully refined and deposited model of the native TLN-**1** (PDB code 5M9W), whereas the superimposed electron density map is the experimentally phased electron density without any further model-based refinement. In the following, the cavities and electron density maps are described relative to the view of Figure 5.2B–E.

Table 5.1: X-ray data collection and refinement statistics.[a]

	Complex (PDB code)				
	TLN-**1** (5M9W)	TLN-**1**-Xe (5M69)	TLN-**1**-Kr (5M5F)	TLN-**5** (5LVD)	TLN-6 (5MA7)
(A) Data collection and processing					
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters: *a*, *b*, *c* (Å)	92.5, 92.5, 130.5	92.5 92.5, 129.9	92.4, 92.4, 130.2	92.8, 92.8, 130.4	92.8, 92.8, 129.8
Wavelength (Å)	0.91841	1.40000	0.85940	0.91841	0.91841
Matthews coefficient (Å^3/Da)[b]	2.3	2.3	2.3	2.4	2.3
Solvent content (%)[b]	47	47	47	48	47
(B) Diffraction data					
Resolution range (Å)	50.00–1.21 (1.28–1.21)	50.00–1.44 (1.53–1.44)	50.00–1.33 (1.41–1.33)	50.00–1.25 (1.32–1.25)	50.00–1.30 (1.38–1.30)
Unique reflections	100158 (15893)	59253 (9235)	75575 (11969)	91830 (14553)	80990 (12879)
R_{sym} (%)	6.2 (48.8)	10.2 (48.6)	8.1 (49.9)	6.6 (49.5)	5.3 (49.0)
Wilson *B* factor (Å^2)	9.9	12.3	10.1	9.4	10.5
Completeness (%)	99.9 (99.5)	99.0 (97.2)	100.0 (99.7)	99.9 (99.3)	99.8 (99.6)
Redundancy	12.9 (12.5)	37.2 (37.1)	20.7 (19.5)	24.9 (24.7)	9.8 (8.7)
$<I/\sigma(I)>$	24.7 (4.9)	28.0 (8.1)	26.0 (6.3)	35.2 (7.0)	26.1 (4.4)
(C) Refinement					
Resolution range (Å)	43.60–1.21	46.24–1.44	43.55–1.33	38.24–1.25	43.67–1.30
Reflections used in refinement (work/free)	95148/5007	56290/2963	71796/3779	87232/4591	76939/4050
Final *R* value for all reflections (work/free) (%)	10.7/12.7	10.8/14.1	11.1/13.9	10.5/12.5	11.2/13.9
FOM	0.95	0.94	0.94	0.95	0.94
Phase error (°)	9.2	11.0	10.6	9.0	10.9
Protein residues	316	316	316	316	316
Calcium/zinc ions	4/1	4/1	4/1	4/1	4/1
Xenon/krypton atoms	0/0	2/0	0/2	0/0	0/0
Inhibitor atoms	28	28	28	30	30
Water molecules	344	332	384	386	380
RMSD from ideality:					
Bond lengths (Å)	0.012	0.013	0.011	0.009	0.010
Bond angles (°)	1.2	1.2	1.2	1.1	1.2
Ramachandran plot:[c]					
Residues in most favored regions (%)	88.9	89.6	87.8	87.8	88.1
Residues in additionally allowed regions (%)	10.0	9.3	11.1	11.1	11.1
Residues in generously allowed regions (%)	0.7	0.7	0.7	0.7	0.4
Residues in disallowed regions (%)[d]	0.4	0.4	0.4	0.4	0.4
Mean *B* factors (Å^2):[e]					
Protein non-hydrogen atoms	11.7	13.8	11.3	11.1	12.3
Inhibitor	14.2	20.5	13.3	12.6	19.3
Water molecules	28.1	30.3	26.2	27.2	28.7

[a] Numbers in brackets stand for the highest resolution shells. [b] Calculated using the program *Matthews_coef* from the *CCP4* suite [137]. [c] Calculated using *PROCHECK* [138]. [d] The Ramachandran outlier Thr26 of TLN is described in the literature [181]. [e] Calculated using *MOLEMAN* [188].

Table 5.2. Zinc MAD dataset of TLN-**1** used for experimental phasing. [a]

	peak	inflection	remote
Space group	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters: *a*, *b*, *c* (Å)	92.5, 92.5, 131.0	92.6, 92.6, 131.0	92.6, 92.6, 131.0
Wavelength (Å)	1.281960	1.283060	0.918410
Resolution range (Å)	50.00–1.50 (1.59–1.50)	50.00–1.50 (1.59–1.50)	50.00-1.32 (1.40-1.32)
Unique reflections	91983 (11525)	91926 (11423)	137883 (17564)
R_{sym} (%)	4.3 (17.2)	3.6 (14.5)	5.4 (46.0)
Completeness (%)	91.4 (71.0)	91.3 (70.3)	93.2 (73.4)
Redundancy	7.8 (7.5)	7.8 (7.5)	7.9 (7.5)
$<I/\sigma(I)>$	31.0 (9.8)	37.0 (12.1)	24.2 (4.2)

[a] Numbers in brackets stand for the highest resolution shells.

The **S_1' cavity** (Figure 5.2B) comprises a volume of 141 Å^3. The top of the cavity is exclusively formed by apolar amino acid side chains of protein residues in addition to the ligand atoms of the P_1' group and the leucine P_2' portion of **1**. The mid- to lower left part of the cavity is mainly apolar, except for Asp138, that, however, is inaccessible for hydrogen bonding due to its buried geometry. The polar side chains of Glu143, Asp170, and Arg203 describe the right-hand surface portion of the cavity. However, except for the side chain of Glu143, they are all highly involved in saturating polar contacts and thus not available for hydrogen bonding with putative occupants of the S_1' pocket. The total electron content integrated over the entire volume of the S_1' cavity is 6.6 e^-.

Reference cavity 1 (Figure 5.2C) is a highly polar cavity comprising a volume of 59 Å^3. It contains three water molecules that establish multiple hydrogen bonds (2.8–3.2 Å) to adjacent polar functional groups of TLN residues. The observed firm density peaks comprise spherical shapes and have been refined to fully occupied water molecules with low *B* factors of 11–12 Å^2. The integrated total electron content within this reference cavity 1 is 19.4 e^-.

Reference cavity 2 (Figure 5.2D) opens a volume of 93 Å^3. In the lower part of the pocket, a strong electron density peak originating from a water molecule (refined *B* factor: 10 Å^2) is detected. This water molecule establishes three hydrogen bonds with Tyr81 (backbone O, 2.9 Å), Arg90 (backbone O, 2.7 Å), and Ser92 (backbone N, 2.8 Å). In contrast, the upper part of the cavity has a highly apolar character. Apart from the described peak assigned to the hydrogen-bonded water molecule, a second, far less intense electron density peak is observed in the center of the large apolar cavity (maximum concentration of 0.62 e^-/Å^3, Figure 5.2D). The distance of this peak to that of the already assigned water molecule is 3.2 Å. The integrated total electron content within reference cavity 2 is 13.4 e^-.

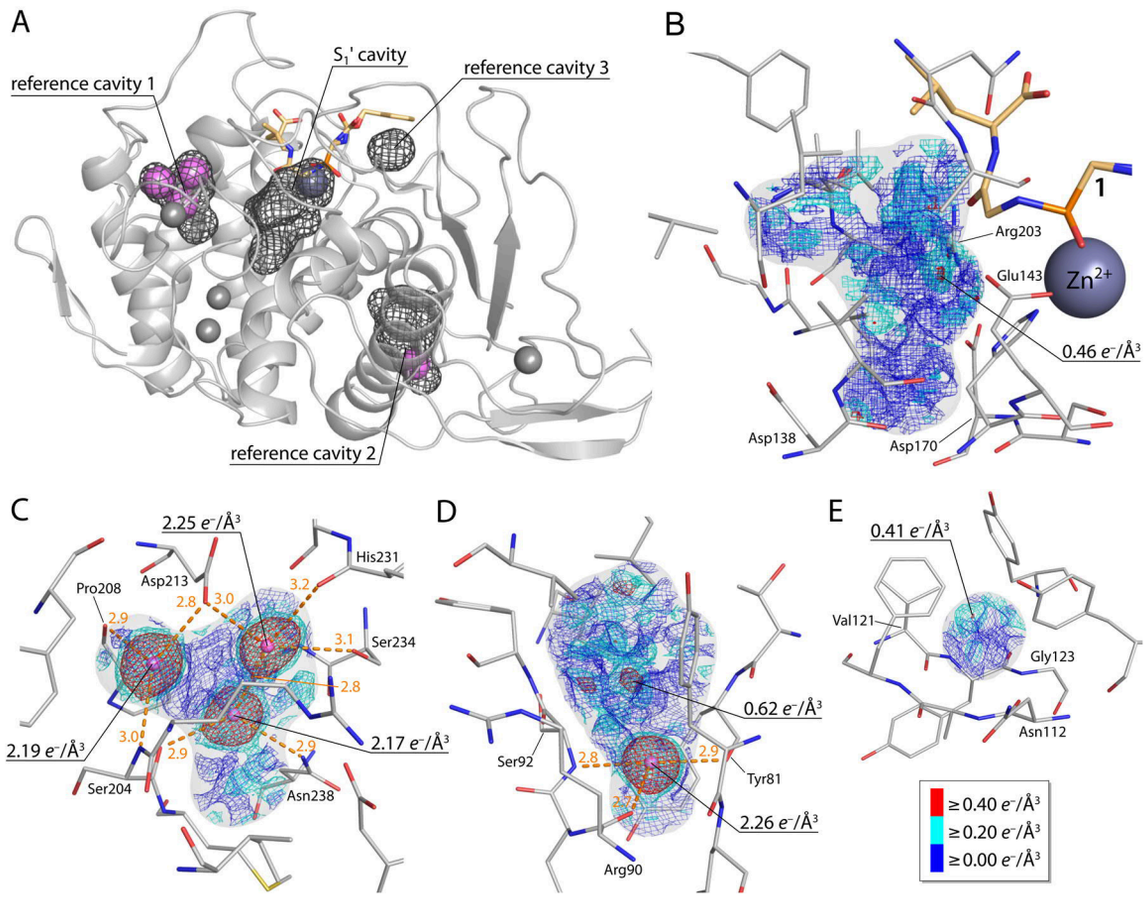


Figure 5.2. Analysis of the experimentally phased electron density map of TLN-**1** within selected cavities. **(A)** Ribbon model of the refined TLN-**1** indicating the solvent excluded surface of the S_1' cavity and of the internal reference cavities 1–3 as dark gray mesh. Water molecules within these cavities are shown as magenta spheres. The zinc ion is shown as dark blue sphere and the four calcium ions of TLN are shown as gray spheres. The TLN-bound ligand **1** is indicated as light orange stick model. **(B–E)** Depiction of the experimentally phased absolute-scale electron density map detected within the **(B)** S_1' cavity (volume 141 Å^3; total electron content: 6.6 e^-), **(C)** reference cavity 1 (volume 59 Å^3; total electron content: 19.4 e^-), **(D)** reference cavity 2 (volume 93 Å^3; total electron content: 13.4 e^-), and **(E)** reference cavity 3 (volume 16 Å^3; total electron content: 0.9 e^-). The electron density map is displayed as blue, cyan or red mesh indicating three different contour levels in e^-/Å^3. Selected density peaks are labeled with their concentration maximum. Cavity lining TLN residues are shown as thin gray stick models. The solvent excluded surfaces of the cavities are indicated in semi-transparent gray. In panel B, the TLN-bound **1** is shown as light orange stick model, and the zinc ion is shown as dark blue sphere. In panels C and D, water molecules are shown as magenta spheres and hydrogen bond interaction distances as orange dashed lines (labeled in Å). The determined volume and total electron content of the cavities are summarized in Table S5.7 (Supporting Information). The electron density maps of all four cavities are shown at six different contour levels from –0.2 to 0.3 e^-/Å^3 in Figure S5.4 to Figure S5.7.

Reference cavity 3 (Figure 5.2E) is, apart from the S_1' cavity, the largest cavity which remains non-solvated in the refined structure of TLN-**1**. Its volume of 16 Å^3 is significantly smaller than that of the S_1' cavity. Its character is primarily apolar, and it is flanked by polar atoms of Asn112 (backbone O), Val121 (backbone O), and Gly123 (backbone N). The integrated total electron content within reference cavity 3 is 0.9 e^-.

5.2.4 Xenon and krypton derivatization of TLN-1

Pressure derivatization of TLN-**1** with xenon and krypton was performed at 9 and 5 bar, respectively. The position of bound xenon and krypton in the crystal structure can be unambiguously identified in the electron density due to their anomalous scattering properties [189]. To optimize the anomalous signal of xenon, a wavelength of 1.4 Å (f' = 6.3 e^-) was chosen, resulting in a strong anomalous signal without significantly losing scattering power due to the long wavelength and air absorption of the X-ray beam (Table 5.1). In the case of krypton, two datasets of the same crystal were collected directly above and below the K absorption edge of krypton to unambiguously identify bound atoms of the noble gas (Table 5.3). Even though the difference between the wavelengths is only about 0.024 Å, the change of the anomalous signal of krypton is drastic, whereas its F_o–F_c electron density virtually remains unaffected (Supporting Information, Figure S5.8). All datasets of the TLN-**1** noble gas derivatives were collected with high redundancy to maximize the accuracy of the anomalous signal (Table 5.1).

Table 5.3. Datasets of noble gas derivatized TLN-**1** used for the determination of the anomalous signal (Friedels pairs separated).[a]

	TLN-**1**-Kr (above *K* edge)	TLN-**1**-Kr (below *K* edge)	TLN-**1**-Xe
Space group	$P6_122$	$P6_122$	$P6_122$
Unit cell parameters: *a*, *b*, *c* (Å)	92.4, 92.4, 130.2	92.5, 92.5, 130.2	92.5, 92.5, 130.0
Wavelength (Å)	0.859400	0.871500	1.40000
Resolution range (Å)	50.00–1.33 (1.41–1.33)	50.00–1.35 (1.43–1.35)	50.00–1.44 (1.53–1.44)
Unique reflections	142833 (23002)	136781 (22091)	111687 (17774)
R_{sym} (%)	7.8 (48.7)	8.2 (46.6)	9.9 (48.0)
Completeness (%)	99.9 (99.7)	99.9 (99.6)	99.2 (97.6)
Redundancy	11.0 (10.2)	8.6 (8.0)	19.8 (19.3)
$<I/\sigma(I)>$	19.2 (4.6)	15.6 (4.0)	20.9 (5.9)

[a] Numbers in brackets stand for the highest resolution shells.

For both noble gases, binding was observed in the S_1' cavity (refined occupancy of xenon: 17%, krypton: 8%) as well as in reference cavity 3 (refined occupancy xenon: 68%, krypton: 20%).

The binding of xenon within the upper, highly apolar part of the deeply buried reference pocket 3 has been previously reported by us (unpublished results, PDB code 3LS7). Both types of noble gases populate at the same positions in the S_1' cavity as well as in the upper part of reference cavity 3. In the S_1' cavity, van der Waals interactions (up to 4.6 Å) are established to the side chains of Val139, His142, Glu143, Ile188, Leu202, Arg203, and the portion of **1** covering the S_1' pocket (Figure 5.3). No adaptations of cavity-lining residues of the noble gas derivatized TLN-**1** are observed compared to the native structure of TLN-**1**.

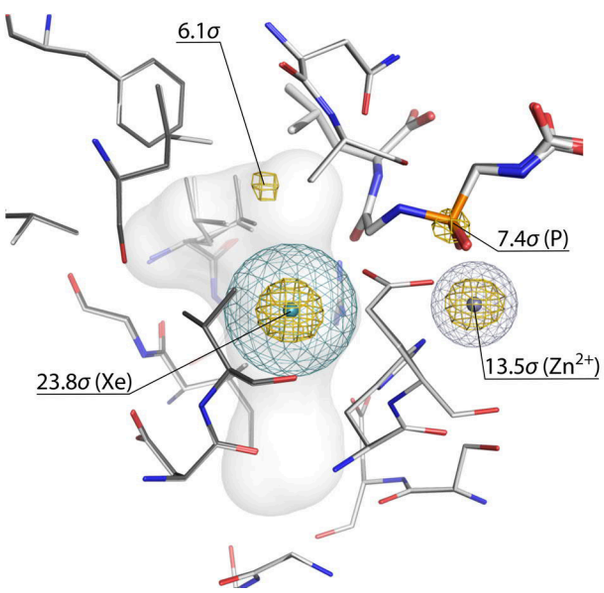


Figure 5.3. Xenon binding site in the S_1' cavity of TLN-**1**. The xenon derivatized crystal structure TLN-**1**-Xe (dark gray) is superimposed on the native crystal structure of TLN-**1** (light gray). The center of the bound xenon atom is shown as teal sphere, and its van der Waals radius is indicated as teal mesh. The zinc ion is shown in similar fashion in dark blue. Cavity surface lining TLN residues are shown as thin sticks, the bound **1** is shown as thick stick model. The solvent excluded surface of the S_1' cavity is indicated in semi-transparent white. The anomalous map is displayed in gold at a contour level of 5σ. The crystal structure of TLN-**1**-Kr is shown in Figure S5.8.

5.2.5 Comparison between the S_1' cavities of TLN-1 to TLN-6

In none of the refined, σ-scaled F_o–F_c electron densities of the six crystal structures TLN-**1** to TLN-**6**, any clearly defined electron density attributable to a bound water molecule could be detected in the S_1' cavity. Even in the complexes of TLN with the ligands exhibiting the polar P_1' groups (**5** and **6**), the unoccupied part of the cavity remains seemingly empty, even though, as indicated in Figure 5.4, sufficient space to accommodate a water molecule is given.

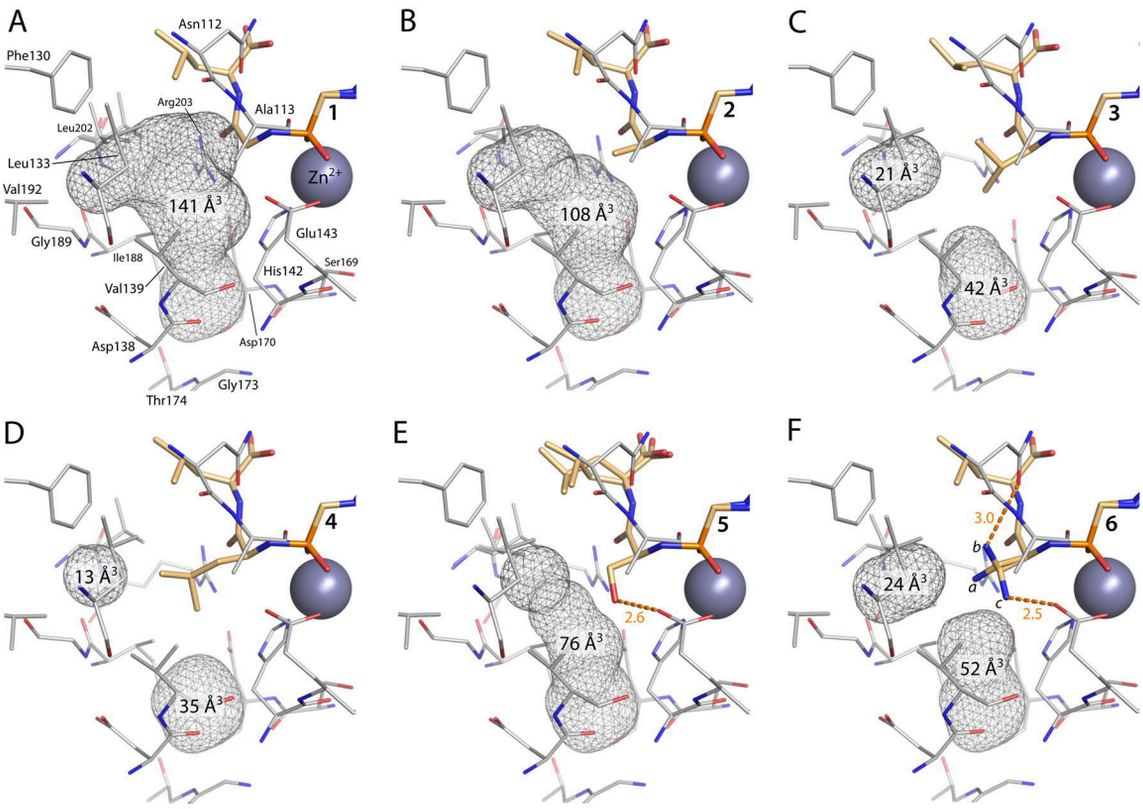


Figure 5.4. S_1' cavities of crystal structures TLN-**1** to TLN-**6**. **(A)** TLN-**1**, **(B)** TLN-**2**, **(C)** TLN-**3**, **(D)** TLN-**4**, **(E)** TLN-**5**, **(F)** TLN-**6**. The solvent excluded surfaces of the S_1' cavities are displayed as dark gray mesh labeled with their volume in Å^3. Cavity lining amino acids are displayed as gray stick models. Ligands **1**–**6** are shown as thick stick models in light orange. The zinc ion is displayed as a dark blue sphere. Hydrogen bonds formed between TLN and the P_1' group of **5** and **6** are indicated as orange dashed lines (labeled in Å). The three conformations of the P_1' group of **6** are labeled with *a*–*c*. All three conformations of the latter P_1' substituent were considered for the calculation of the residual cavity volume.

Figure 5.4 displays the crystal structure models of TLN-**1** to TLN-**6** and the residual S_1' cavities. With growing P_1' portion (**1**→**4**), the top part of the cavity is gradually occupied, until only the bottom part of the cavity remains unoccupied (Figure 5.4A–D). Accordingly, the volume of the S_1' cavity is gradually decreasing from 141 Å^3 to 48 Å^3 (Figure 5.4). The hydrophobic P_1' groups of TLN-**2**, TLN-**3** and TLN-**4** form multiple hydrophobic van der Waals interactions to the pocket-shape determining amino acid side chains of TLN. Whereas the P_1' OH group of the serine derivative is found in only one populated conformation forming a hydrogen bond to the side chain of Glu143 (2.6 Å, Figure 5.4E), three conformations (occupancy *a*: 42%, *b*: 27%, and *c*: 31%) can be assigned to the P_1' amino function of **6** in the refined model (Figure 5.4F). Conformations *b* and *c* place the amino group within hydrogen-bonding distance to the side chain of Asn112 (3.0 Å) and Glu143 (2.5 Å).

5.3 Discussion

5.3.1 Analysis of the experimentally phased electron density within the S_1' cavity of TLN-1

Even with high-resolution data as in the current case, the analysis of the solvation pattern within a protein cavity by X-ray crystallography is by no means straightforward, since highly mobile ("disordered") or partially occupied water molecules are difficult to detect on first glance. A highly mobile water molecule may be characterized by the lack of a well-defined, sufficiently deep energy minimum on the free energy landscape, resulting in a widely distributed, rather blurred electron density lacking a clearly defined center [190]. Model bias introduced into the electron density via the use of phases transferred from the refined structure of a related complex can obscure the detection of such weak density signals. In particular, the use of the commonly applied molecular replacement as phasing technique can result in a significant impact of model bias [191–193]. Experimentally phased electron densities have the advantage that no model bias is arbitrarily introduced by the application of predefined model phases [187,194,195]. Therefore, to reliably detect traces of electron density originating from highly mobile or partly occupied water molecules, we performed an elaborate experimental phasing protocol for TLN-**1**. Even though unbiased phases can be obtained by this procedure, another obstacle arises from the fact that the total electron number within the crystal unit cell F_{000} is impossible to determine experimentally in the case of proteins. In small molecule crystallography, the content of the diffracting unit cell is usually easy to define and thus the number of contributing electrons is clear. In the case of proteins showing a large solvent content particular in the channels passing through the crystal, such an assignment is impossible and thus can only be estimated. In consequence, electron densities in protein crystallography are typically σ-scaled [196], where zero σ corresponds to the average, numerically unknown electron concentration of the unit cell. Even though this step appears very reasonable and pragmatic in usual standard refinements of protein structures, it will make an assignment of an absolute electron number of electrons to a particular integrated volume virtually impossible. Therefore, we attempted to transform the experimentally phased electron density to an absolute electron number scale, where the zero value corresponds to vacuum. Following a similar approach, Liu et al. analyzed a hydrophobic cavity of 134 $Å^3$ in L99A/M102L T4 lysozyme [195]. In this case, analysis of the experimentally phased electron density on absolute scale discovered a water cluster of three water molecules with an occupancy of approximately 50% within this pocket. In another example, the analysis of an experimentally phased electron density of a central apolar cavity of 40 $Å^3$ observed in interleukin-1β settled the debate about the presence of a disordered water molecule in this volume [194,197–199]. From the data analysis, Quillin et al. concluded that the cavity must indeed be empty [194].

Reference cavity 3 is considered as an internal control for an entirely empty cavity (Figure 5.2E). This small, apolar cavity of 16 Å^3 is due to its size and polarity unsuited to host a water molecule [200–202]. With respect to the calculated volume of 16 Å^3 for reference pocket 3 it must be considered that the calculation of the volume that is really available to host a water molecule is non-trivial and is actually further reduced adjacent to apolar residues [203,204]. In other words, the solvent probe radius adjacent to apolar residues is in fact larger than the commonly applied 1.40 Å. Consequently, the presence of a water molecule in reference cavity 3 can be certainly excluded. The electron concentration peak found within this cavity comprises a concentration maximum of 0.41 e^-/Å^3, and the total electron content is 0.9 e^-. We consider this electron content the result of "spillover" of electrons from cavity lining amino acids into the to some degree fuzzy volume of the empty cavity [194,195], e.g. due to a slight movement of the protein residues. Reference cavities 1 and 2 (Figure 5.2C and D) serve as an internal control for clearly solvated cavities. The calculated total electron content within reference cavity 1 (59 Å^3) is 19.4 e^-, a number significantly lower than expected for three fully occupied water molecules (30 e^-). Similar as observed by Liu et al. [195], this is possibly the result of the electron spillover from within the cavity to the outside due to overlapping van der Waals radii of polar cavity-lining atoms and the tightly hydrogen-bonded water molecules. Reference cavity 2 (93 Å^3) contains a total electron content of 13.4 e^-. Apart from the unambiguous water density in the lower part of the cavity, a significant amount of blurred electron density with a peak maximum of 0.62 e^-/Å^3 is detected in the balloon-like shaped upper, hydrophobic part of the cavity (Figure 5.2D) — which is large enough to host a phenol molecule without that the water molecule in the lower part of the cavity is displaced [205]. The electron concentration maximum is within hydrogen bonding distance (3.2 Å) to the modeled water molecule in the lower part of the cavity and could potentially originate from a second, very low populated water site. Overall, this results in a higher electron content than expected for the single fully occupied water molecule, overcompensating the also in this case expected spillover of electrons of the tightly bound water molecule to the outside of the cavity. Remarkably, the residual electron distribution within the S_1' cavity shows an electron concentration of maximally 0.46 e^-/Å^3, a value close to the one found for the empty reference cavity 3 (0.41 e^-/Å^3). The total electron content within the S_1' cavity was determined to be 6.6 e^-. If the electron content within the empty reference cavity 3 (0.9 e^-) is extrapolated to the larger volume of the S_1' cavity (141Å^3/16 Å^3 = 8.8 times larger), this corresponds to 7.9 e^- (0.9 e^- × 8.8), a number higher than the one for the experimentally determined electron content within the S_1' cavity (6.6 e^-). This can be explained by the fact that relative to their volumes reference cavity 3 has a significant larger surface than the S_1' cavity, resulting in a proportionally larger electron "spillover" (*vide supra*). Taking into account these considerations we therefore propose that the TLN-**1** S_1' cavity is indeed a completely unsolvated cavity. We also believe that this is not simply a consequence of the binding of **1**

which seals the pocket from the top. Our complex structures of TLN-**1** with the noble gases clearly demonstrate that the pocket is still well accessible in the crystal by particles as large as a xenon atom, likely due to the residual mobility of the complexes. Thus, smaller water molecules could easily access the S_1' pocket if their penetration would be favorable.

5.3.2 Analysis of TLN-1 in complex with xenon and krypton and of crystal structures TLN-1 to TLN-6

The noble gases xenon and krypton are known to preferentially bind to desolvated, hydrophobic protein cavities through weak van der Waals interactions [206,207]. These atoms can, therefore, be used as experimental probes to detect such cavities. The S_1' cavity (Figure 5.3) and reference cavity 3 were revealed as noble gas binding sites in TLN-**1**. The fact that both noble gases were found binding to the S_1' cavity even though the derivatization pressure was kept low supports the hypothesis that both cavities do not contain significantly populated solvent molecules, that need to be displaced upon ligand accommodation. The increased occupancy of the gaseous probes in reference cavity 3 compared to the S_1' cavity can be attributed to the deeper burial of this cavity within the apolar interior of TLN. Thus, during the experimental depressurization phase, that needs to be accomplished to transfer the crystal specimen from the pressuring cell to the liquid nitrogen, diffusion of the noble gases is slower. The reduced occupancy of krypton compared to xenon is attributable to the reduced polarizability of the former [208,209], the lower applied derivatization pressure, and the faster diffusion kinetics of the smaller krypton atoms. Noble gas binding provides a crude estimate about the at least available space within a given cavity, and provides information about an at least detectable electron concentration in a certain volume [208,210]. For instance, in TLN-**1**-Kr (Figure S5.8), the krypton atom (36 e^-, van der Waals radius 2.0 Å, volume 33.5 Å^3) bound to the S_1' cavity was refined to 8% occupancy. This corresponds to 2.9 electrons (36 $e^- \times 0.08$). Considering the refined temperature factor of 12.5 Å^2 of this krypton atom, krypton will occupy a volume of about 58 Å^3 [190,210]. Hence, if we transfer this detection limit to our cavity analysis, a putative water molecule occupied as low as 29% (corresponding to 2.9 e^-) and distributed over a volume of 58 Å^3 should be detectable by conventional refinement. If we apply this estimation rule to the S_1' cavities of complexes TLN-**1** to TLN-**6**, it is permissible to conclude that the residual volumes found in TLN-**3** to TLN-**6** (48–76 Å^3, Figure 5.4) must be virtually empty. This is further supported by the observation that the S_1' cavities of TLN-**3**, TLN-**4** and TLN-**6** are divided by the P_1' groups of the bound ligands into two spatially separated cavities, reducing the mobility of a putatively present water molecule even further. Consequently, a potentially present — but nonetheless undetectable — water molecule must be very low occupied to elude crystallographic detection completely. We believe that this minor displacement effect would be insignificant for the binding event and would hardly influence

the overall thermodynamic profile. Thus, for the discussion of the thermodynamic signature, we refer to the empty state of the cavity.

The polar groups of **5** and **6** were introduced into the S_1' pocket with the aim to provide a local anchor for hydrogen bonding for a potentially present and disordered water molecule. If such a water molecule would hypothetically be present in the complexes with the aliphatic ligands, this hydrogen-bonding decoy should result in the stabilization of a disordered water molecule — or possibly in the recruitment of a new water molecule from the bulk water phase — and make it crystallographically detectable [211,212]. However, no clear electron density peak is observed in the crystal structures of TLN-**5** or TLN-**6**. One reason to be regarded why the polar groups are not available to establish hydrogen bonds to a putative water molecule might arise from the fact that they form hydrogen bonds to protein residues. However, conformation *a* with the highest population of the amino function of **6** orients toward the void of the cavity where it could experience hydrogen bonding (Figure 5.4F). Another reason could be that even though the polarity of the S_1' cavity of TLN-**5** and TLN-**6** increases, the volume of the cavity is reduced, which decreases the probability to find a water molecule within a cavity [201].

5.3.3 Thermodynamic binding profiles of 1–6 as determined by ITC

The affinities of the investigated ligand series fall into the range between milli- to nanomolar binding (Table S5.1). Because of this broad range, it was necessary to apply different measurement protocols to obtain reliable calorimetric data. Ligands **2**, **3** and **4** were measured by direct titration. Due to their low affinities, **1**, **5** and **6** had to be characterized by displacement titrations. The measurement accuracy of the latter ligands is strongly dependent on how accurately the thermodynamic data of the applied reference ligand has been recorded. Any error in the thermodynamic profile of the reference ligand will propagate to the thermodynamic profile of the analyte. Consequently, such measurements usually result in large experimental errors, in particular considering the partitioning of enthalpy and entropy [96]. The affinity of **1** is too low to accurately determine the enthalpy/entropy signature via displacement titration. Furthermore, **5** and **6**, comprising polar P_1' groups, show very likely different changes in their protonation state compared to the ligands with the aliphatic P_1' groups. A superposition of a proton transfer reaction can alter the observed enthalpy of binding and therefore prevent a direct comparability of measured enthalpy values [213]. Unfortunately, an appropriate enthalpy correction is not reliably feasible in a displacement titration scenario due to the resulting highly potentiating errors of the buffer-corrected thermodynamic profiles. Consequently, we decided to refrain from analyzing the enthalpy/entropy profiles of **1**, **5** and **6**, and instead solely report their affinities. This value is independent of buffer effects and allows accurate comparison with the more potent ligands.

The addition of a single methyl group to the glycine derivative **1** (yielding **2**) results in a more than 100-fold affinity increase (Figure 5.1B and Table S5.1). Adding two further methyl groups to **2** (yielding **3**) results in a further 40-fold affinity increase. Finally, the addition of a further methyl group from **3** to **4** corresponds to a 7-fold affinity increase. Based on a statistical evaluation, a 100-fold affinity increase solely achieved by the addition of one single methyl group has been reported to be a very rare event with a probability of only 0.4% [214]. In the reported cases, the favorable affinity increase resulted from the placement of a methyl group into hydrophobic pockets that the authors assumed to be entirely empty, in combination with an energetically favorable pre-organization of the ligand conformation in solution. This assumption is supported by MD simulations suggesting that the placement of hydrophobic groups into supposedly empty, hydrophobic pockets produces an extraordinarily favorable change of the Gibbs free energy of binding due to the absence of a cavity desolvation step [215,216]. In the same manner, the extraordinary affinity increase from **1** to **2** (>100-fold) can be attributed to the insertion of a methyl group into a hydrophobic cavity, that is — as experiment confirms — virtually empty. This effect is reduced, but nevertheless still pronounced when further methyl groups are grown into this volume (40-fold affinity increase between **2** and **3**). From **3** to **4**, the comparably low 7-fold affinity increase is the result of a highly favorable enthalpic term, that, however, is partly compensated by unfavorable entropy (Figure 5.1B). Usually, dispersive (van der Waals) interactions are less significant, because such interactions established between protein and ligand are largely canceled out by the required unfavorable disruption of van der Waals interactions between protein and solvent [217]. However, in the case of a vacated pocket that makes a desolvation step obsolete, dispersive interactions formed between protein and ligand upon complexation become determinant in terms of affinity [217]. This is the case in the current study — there is no cost to desolvate the S_1' binding pocket. As a result, the contribution of the established dispersive interactions between protein and ligand to the enthalpy of binding increases with increasing P_1' chain length, and overall strongly affects binding affinity. However, with increasing P_1' chain length, the enthalpic signal is increasingly compensated by a decrease in entropy (Figure 5.1B). This is potentially the result of a loss of conformational degrees of freedom upon complex formation. The low affinities of **5** and **6** can be attributed to the large energetic penalty for the desolvation of their polar functional groups which is not overcompensated by the dispersive interactions with the cavity-lining residues. The hydrogen-bond interactions formed with Glu143 (and Asn112 in the case of **6**) do not suffice to compensate for this loss. Thus, the free energy gain resulting from the establishment of dispersive interactions is largely compensated by the high cost to desolvate **5** and **6**, overall lowering their affinity.

Protein cavities can fulfill essential biological functions, e.g. conformational flexibility, and thus can represent more than "packing defects" [207,218,219]. In the case of TLN (and many other metalloproteinases), the S_1' pocket is the most important pocket to discriminate

substrates from non-substrates, and thus defines the selectivity profile of the protease. A substrate that exhibits shape complementary and thus fills the hydrophobic S_1' pocket efficiently without requiring a large price for desolvation (e.g. a P_1' leucine, *iso*-leucine or valine side chain) will experience strong dispersive interactions with the protease. It thus gets bound and enzymatically processed. In contrast, if the pre-organized void in the protease is insufficiently filled by an either small or hydrophilic P_1' substrate portion, a pronounced affinity reducing enthalpic penalty results. Thus, the unsolvated state of the TLN S_1' pocket is highly important for the selectivity mechanism of the protease.

An entirely independent proof for the existence of unsolvated protein binding pockets is provided by a current neutron diffraction study [220]. Neutron scattering can reliably differentiate between hydrogen and deuterium. We recently determined the joint X-ray/neutron structures of two trypsin complexes. Crystal growth was performed with the protein in its hydrogen form and crystals were subsequently exposed to fully deuterated buffer over 532 days before data collection. Under such conditions polar hydrogen atoms should exchange by deuterium atoms considering the large excess of deuterons compared to protons. However, as a precondition the polar groups must be accessible to D_2O molecules. In a folded protein, polar hydrogens usually remain only at sites where they are involved in strong hydrogen bonds. Nevertheless, in a sterically accessible pocket of our trypsin complexes we found NH groups where the polar hydrogens are not involved in a hydrogen bond and these NH groups did hardly exchange to ND over the 1.5 years of incubation in D_2O. Only 7% deuterium could be found at these sites. We therefore hypothesize that they hardly experienced any contact with D_2O molecules in the folded protein indicating that these cavities are extremely rarely accessed by water molecules. We believe this is another independent indication that empty and hardly unsolvated pockets exist in folded proteins.

5.4 Conclusion

The discussed experimental data strongly suggest that the S_1' cavity of TLN in complex with ligands **1**–**6** is not solvated and contains virtually vacuum prior to ligand binding. We found no experimental evidence for the occupation with highly mobile water molecules, and the properties of the cavities (shape, volume, polarity) suggest that the cavity is indeed empty. It seems energetically more favorable to maintain a vacuum than to host one or several water molecules within these cavities. It has been discussed that the generation of vacated pockets in proteins is energetically very costly [221,222]. Nevertheless, it has to be considered that the costs for producing such a void have to be afforded during protein-folding and not during ligand or substrate binding.

In the current case, the fact that the TLN S_1' cavity is empty takes major consequences on the energetically highly favored accommodation of aliphatic P_1' side chains of either substrate or inhibitor molecules. The observed enthalpy driven affinity enhancement with increasing size of the P_1' substituent (**1**→**4**) is mainly a result of the binding of aliphatic groups into a void, where no price for pocket desolvation has to be afforded. The decreasing affinity contribution of a growing side chain can be attributed to the augmenting desolvation penalty of the larger P_1' substituents and to a reduction in conformational flexibility. Addition of a polar group entirely destroys binding affinity due to an uncompensated desolvation penalty. Remarkably, the derivative with a P_1' benzyl side chain (PDB code 3FV4) that fills the S_1' pocket more efficiently than the P_1' leucine side chain of **4** is less potent, likely due to the higher desolvation costs for the aromatic side chain [33]. In the case of TLN, the hydration state of the S_1' cavity seems to have a pronounced contribution to the selectivity profile of the protease. The remarkable, about 41,000-fold increase in affinity from the introduction of hydrophobic *iso*-butyl group into this pocket also has major implications for medicinal chemistry. The identification of poorly hydrated cavities can represent a valuable strategy to gain overwhelmingly in binding affinity of a prospective drug molecule.

5.5 Experimental section

5.5.1 Ligand synthesis

The synthesis of **1**–**6** has been reported previously [183].

5.5.2 Crystal preparation and soaking

Crystals were prepared with lyophilized TLN powder commercially obtained from Calbiochem (EMD Biosciences) identical to the procedure as described earlier [16]. For crystal soaking of the low-affinity compounds **1**, **5** and **6**, TLN crystals were transferred into a soaking solution composed of 100 mM Tris-HCl, pH 7.5, 2 mM $CaCl_2$ and 20% DMSO saturated with the respective ligand (ligand precipitate visible), followed by incubation for 24 hrs. Afterward, crystals were flash-frozen in liquid nitrogen after a brief immersion in a cryoprotectant solution saturated with the respective ligand, composed of 10 mM Tris-HCl, pH 7.5, 10 mM $Ca(CH_3COO)_2$, 20%DMSO, and 20% glycerol.

5.5.3 Derivatization of TLN with xenon and krypton

For the noble gas derivatization of the TLN crystals (TLN-**1**-Xe and TLN-**1**-Kr), a pressure cell from Oxford Cryosystems (Long Hanborough, UK) was used [223]. Before derivatization, TLN crystals were soaked with **1** in the above-mentioned soaking buffer for 24 hrs. To protect the crystals from drying out during the pressurization phase, the filter paper of the pressurization cell was drenched with soaking buffer. Subsequently, xenon derivatization was performed at 9 bar for 5 minutes, and derivatization with krypton was carried out at 5 bar for 5 minutes. Derivatization was conducted at relatively low pressure because higher pressure resulted in a strong increase in crystal mosaicity. After the incubation time was up, pressure was quickly released, and TLN crystals were immediately after flash-frozen in liquid nitrogen.

5.5.4 Data collection

Datasets for crystal structures TLN-**1**, TLN-**5**, and TLN-**6** (Table 5.1) were collected at BESSY II (Berlin-Adlershof, Germany) on beamline 14.1 on a Dectris Pilatus 6M pixel detector at 100 K [157,224]. A three-wavelength MAD dataset of TLN-**1** (Table 5.2) which was applied for the determination of the experimental phases was collected from another crystal at BESSY II on beamline 14.2 on a Rayonix MX-225 pixel detector at 100 K. Datasets for crystal structures TLN-**1**-Xe and TLN-**1**-Kr (Table 5.1) were collected at Elettra (Triest, Italy) on beamline XRD1 at 100 K on a Dectris Pilatus 2M pixel detector. The dataset for TLN-**1**-Xe was collected at λ =

1.40000 Å. Two datasets were collected for TLN-**1**-Kr from the same crystal (Table 5.3), the first at 0.85940 Å above the krypton *K* absorption edge (that is at 0.86550 Å), against which the structure was refined, and a second dataset at 0.871500 Å below the krypton *K* absorption edge. The second dataset was solely used for analysis of the change of the anomalous signal of krypton to unambiguously identify the bound atom. The anomalous maps TLN-**1**-Xe (Figure 5.3) and TLN-**1**-Kr (Figure S5.8) were created with *ANODE* [225].

5.5.5 Dataset processing and structure determination

All data sets were indexed, integrated and scaled with *XDS* [158]. The phases of TLN-**1**-Xe, TLN-**1**-Kr, TLN-**5**, and TLN-**6** were determined by molecular replacement applying *Phaser* [159] from the *CCP4* suite (version 6.3.0) [137]. The crystal structure of the PDB entry 8TLN was used as molecular replacement search model [160]. The phases for TLN-**1** were determined experimentally by a MAD dataset (Table 5.2) using the intrinsically bound zinc ion of TLN as anomalous scatterer. Experimental phasing was performed applying the programs *SHELXC* (data preparation), *SHELXD* (heavy atom substructure determination) [226] and *SHELXE* (experimental phasing and density modification) [227] as implemented in *HKL2MAP* [228]. The anomalous signal d''/σ of the peak dataset was significant (1.42) to 1.62 Å. Between all datasets the anomalous correlation coefficient was above 30% up to 1.57 Å. The best solution of the *SHELXD* substructure search was CC_{All}=71.3 and CC_{Weak}= 58.9. The substructure phases calculated by *SHELXE* from the original hand gave a final contrast of 0.56 and a connectivity of 0.81 (inverted hand: contrast 0.35, connectivity 0.66). The experimentally determined phases were combined with the amplitudes of a 1.21 Å native dataset of TLN-**1** (Table 5.1) followed by density modification with the program *DM* [229] from the *CCP4* suite applying solvent flattening and histogram matching. A starting model for conventional refinement was created with *ARP/wARP* [230], where 314 amino acids in a single polypeptide chains (99% sequence coverage) were successfully placed into the electron density map with a resolution of 1.21 Å.

5.5.6 Model building and refinement

Crystal structure model refinement was performed with *phenix.refine* version 1.10.1-2155 [162]. Simulated annealing with default settings was performed as first refinement step. Subsequently, all crystal structure models were refined with riding hydrogen atoms added to protein residues, applying xyz refinement, individual anisotropic *B* factors for all atoms except for hydrogens, and occupancy refinement. Refinement cycles were intermittent by model building with *Coot* [161]. Ligand building was performed with *MOE* [151], ligand restraints

were created with *eLBOW* [163]. Graphical representations of the crystal structure coordinates and electron density maps (automatic ccp4 map normalization turned off) were created with *PyMOL* [164].

5.5.7 Cavity computation and volume calculation

The cavities as displayed in Figure 5.4 were computed with *POVME* [231]. Dummy atoms (DAs) with a radius of 1.40 Å were placed into the cavities (grid spacing 0.2 Å). The radius of 1.40 Å was chosen as it is frequently applied as the radius of a water molecule in the literature [194,208,210,232,233]. The solvent excluded surfaces of the DA objects representing the cavities were then displayed with *PyMOL*. Since *POVME* can only calculate the solvent accessible volume, the solvent excluded volumes of the DA objects describing the S_1' cavities of TLN-**1** to TLN-**6** (Figure 5.4) were calculated with the program *3V* (radius of the DAs set to 1.40 Å) [234]. The volumes of the S_1' cavity and of reference cavities 1–3 as displayed in Figure 5.2 used for the analysis of the electron density map were created with DA comprising a radius of 1.30 Å. The reduction of the DA radius compared to the DA radius applied for computing the cavities in Figure 5.4 (1.40 Å) was necessary, otherwise the narrow parts of reference cavities 1 and 2, where hydrogen-bonded water molecules are closely bound to TLN residues, would not have been detected.

5.5.8 Placement of the experimentally phased electron density map of TLN-1 on an absolute electron number scale and determination of the total electron content within a cavity

Two mathematical operations are required to transform the experimentally phased, σ-scaled map of TLN-**1** (ccp4 format) characterized by arbitrarily small map voxel values to an absolute electron number density map where zero corresponds to vacuum — thus allowing quantification of the total electron content within a given map volume. First, the values of the map voxels have to be set to the correct scale by applying a scaling coefficient. Subsequently, the still σ-scaled map — where zero corresponds to the average electron concentration of the unit cell — has to be shifted in a way that zero corresponds to vacuum. For that, the average electron concentration ($e^-/Å^3$) of the unit cell must be added to every map voxel. To have access to an absolute electron number density map as a reference, the σ-scaled density map of the fully refined model of TLN-**1** was transformed to an absolute electron number density map applying the *END* map script developed by Lang et al. [235] running with *phenix.refine* version 1.8.4-1492. This program computes absolute *electron number density* (*END*) maps from conventionally refined σ-scaled maps. However, it is dependent on a refinement program and

relies on refined model phases for the calculation of the electron density map, thus it is not possible to directly use it to transform the experimentally phased σ-scaled density map to an absolute electron number scale without the introduction of model phases. The *END* map script also provided the structure factor F_{000} of TLN-**1** (256668 e^-), which corresponds to the total electron content within the crystal unit cell under consideration of protein model and bulk (disordered) solvent. Thereby, the average electron concentration within the unit cell was accessible by F_{000}/unit cell volume = 256668 e^-/967353 Å^3 = 0.26533 e^-/Å^3.

We decided to use the Zn^{2+} ion (28 e^-) that is intrinsically bound to TLN to derive the scale coefficient. To be independent of the theoretical electron number of the zinc ion and of the necessity of an integration mask covering the entire zinc ion without any electron spillover from inside to the outside of the mask (e.g. due to a slightly disordered zinc) or electron spillover from outside to the inside of the integration mask (e.g. from zinc-complexing residues), we determined the total electron content of a spherical map fragment describing the zinc ion of the refined *END* map. The zinc coordinates of the refined crystal structure TLN-**1** were used for the center of the sphere, and 1.39 Å was used as sphere radius. The total electron content within this map fragment gave 27.3 e^- as calculated with an *in-house* script based on the program *mapman* [236]. Subsequently, an identical integration mask was applied to the σ-scaled experimentally phased electron density for scaling of the latter dataset by applying the program *mapmask* from the *CCP4* suite [137]. The scale coefficient (14.1) was derived in a way that after applying the scale coefficient followed by addition of the average electron concentration (0.26533 e^-/Å^3), integration of the map fragment covering the zinc of the experimentally phased map resulted in exactly 27.3 e^-, just as observed in the map fragment of the refined *END* map. Visually, the experimentally phased electron density map was correctly shifted in a way that zero corresponds to vacuum (Figure S5.4 to Figure S5.7). If an electron concentration of ≥0.0 e^-/Å^3 is displayed (panels C of Figure S5.4 to Figure S5.7), the entire cavities are covered with electron density. In contrast, displaying higher (positive) or lower (negative) electron density concentrations for instance within the empty S_1' pocket (Figure S5.4) immediately reduces the visible residual electron density.

For the calculation of the total electron content within a given volume, e.g. the volume of the S_1' cavity and reference cavities 1–3, a map fragment was cut out from the entire electron density map applying the program *Coot* via the command "mask map by atom selection", followed by integration over all voxels (including negative values) applying the aforementioned *in-house* script. For the S_1' cavity and reference cavities 1-3, the DA objects also used for displaying the contour regions of the cavities as shown in Figure 5.2 were used applying a radius of 1.3 Å.

5.5.9 Measurement of the thermodynamic binding profiles

For the ITC measurements, a Microcal ITC_{200} device from GE Healthcare (Piscataway, New Jersey) was used. Measurements were performed with freeze-dried TLN powder bought from Calbiochem (EMD Biosciences), which was dissolved in buffer directly before measurement without further treatment. Measurement buffers (pH 7.5) were composed of 20 mM ACES, MOPSO, MES, PIPES or cacodylate buffer substance, 500 mM NaSCN, and 2 mM $CaCl_2$. The salt NaSCN was selected because it strongly increases the solubility of TLN [237]. The high concentration of 500 mM was necessary to provide solubility of TLN up to 250 µM that was necessary to enable a direct measurement of the low-affinity ligand **2**. Since different concentrations of NaSCN are known to influence the measured thermodynamic binding parameters [96], this salt concentration was applied for all ligands to guarantee relative comparability of the thermodynamic binding parameters. The thermodynamic profiles of **4** (Table S5.4) and **3** (Table S5.5) were determined by direct titration applying a TLN concentration of 50 µM, resulting in well analyzable isotherms with clear sigmoidal curvatures (Figure S5.1C and B). For the low-affinity compound **2** (Table S5.3), the concentration of TLN was increased to 250 µM, resulting in isotherms described by a *c*-value of 6 (Figure S5.1A). This allowed the experimental determination of the inflection point and the determination of reliable thermodynamic parameters. To determine the heat of ionization associated with the complex formation of TLN and **2**, **3** or **4** (Figure S5.2), measurements were performed in ACES, MOPSO, HEPES, MES, PIPES and cacodylate buffers. The affinities of the weak binding compounds **1**, **5** and **6** were determined by displacement titrations (Figure S5.2D–F and Table S5.6) [238]. 50 µM TLN cacodylate buffer solution was preincubated with different concentrations of weak ligand (**1**: 5, 8, 20 mM; **5**: 5, 10, 15 mM; **6**: 1.5, 2, 2 mM), followed by titration with the reference ligand **4**. Different concentrations of the weak ligand (resulting in titration curves exhibiting different *c*-values) were applied to improve the accuracy of the binding parameter determination by global analysis of the ITC isotherms [184]. Measurement peaks of the raw thermograms were extracted and integrated with *NITPIC* version 1.1.2 [167,239]. Global analysis of the ITC isotherms was performed with *SEDPHAT* version 12.1b [168,240]. The thermodynamic profiles of the ligands measured in individual buffers were determined by global analysis applying the model "A+B ↔ AB Hetero-Association". Correction for the heat of ionization of these thermodynamic profiles was performed using the model "A+B ↔ AB Hetero-Association Global Buffer Ionization Enthalpy Analysis". For the displacement titrations, the model "A+B+C ↔ AB+C ↔ AC+B; competing B and C for A" was applied. ITC isotherm graphs were prepared with *GUSSI* [169].

5.6 Abbreviations

TLN, thermolysin; ITC, isothermal titration calorimetry; FOM, figure of merit; DA, dummy atom; ACES, *N*-(2-acetamido)-2-aminoethanesulfonic acid ; MOPSO, 3-morpholino-2-hydroxypropanesulfonic acid; PIPES, piperazine-*N*,*N*-bis(2-ethanesulfonic acid)

5.7 PDB accession codes

Newly reported crystal structures are available in the PDB upon publication of this manuscript (accession codes in brackets): TLN-**1** (5M9W), TLN-**5** (5LVD), TLN-**6** (5MA7), TLN-**1**-Xe (5M69), and TLN-**1**-Kr (5M5F). The crystal structures of TLN-**2** (3FV4), TLN-**3** (3FLF), and TLN-**4** (4H57) were previously reported [33].

5.8 Acknowledgements

S. G. K. and J. C. contributed equally to this work. The authors want to thank the MX beamline teams at BESSY II (Berlin-Adlershof, Germany) and Elettra (Triest, Italy) for advice during data collection. In particular, the authors want to thank Dr. Maurizio Polentarutti from Elettra Sincrotrone for help with the noble gas derivatization experiments. The receipt of a travel grant from the Helmholtz-Zentrum Berlin is gratefully acknowledged. This work was supported by grant 268145-DrugProfilBind awarded to G. K. by the European Research Council (ERC) of the European Union.

5.9 Supporting Information

5.9.1 Isothermal titration calorimetry results

Table S5.1 (related to Figure 5.1). Thermodynamic binding profiles determined for **1**–**6**. The values in parenthesis describe the boundaries of the 95.4% confidence interval. All values are rounded. The thermodynamic profiles of **2**, **3** and **4** were corrected for the heat signal of buffer ionization upon complex formation.

ligand	K_d (μM)	$\Delta G°$ (kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})
1[a]	5,659 (8,321, 3,762)	-12.8 (-11.9, -13.8)	-	-
2[b]	40.750 (45.793, 36.299)	-25.1 (-24.8, -25.3)	-22.7 (-23.6, -21.9)	-2.4 (-1.2, -3.4)
3[c]	0.992 (1.234, 0.792)	-34.3 (-33.7, -34.8)	-33.7 (-34.8, -32.7)	-0.5 (1.1, -2.1)
4[c]	0.138 (0.201, 0.090)	-39.1 (-38.2, -40.2)	-45.1 (-46.5, -43.7)	5.9 (8.3, 3.5)
5[a]	2,407 (3,652, 1,507)	-14.9 (-13.9, -16.1)	-	-
6[a]	336 (752, 168)	-19.8 (-17.8, -21.5)	-	-

[a] Determined by displacement titration. [b] Direct titration with 250 μM TLN. [c] Direct titration with 50 μM TLN.

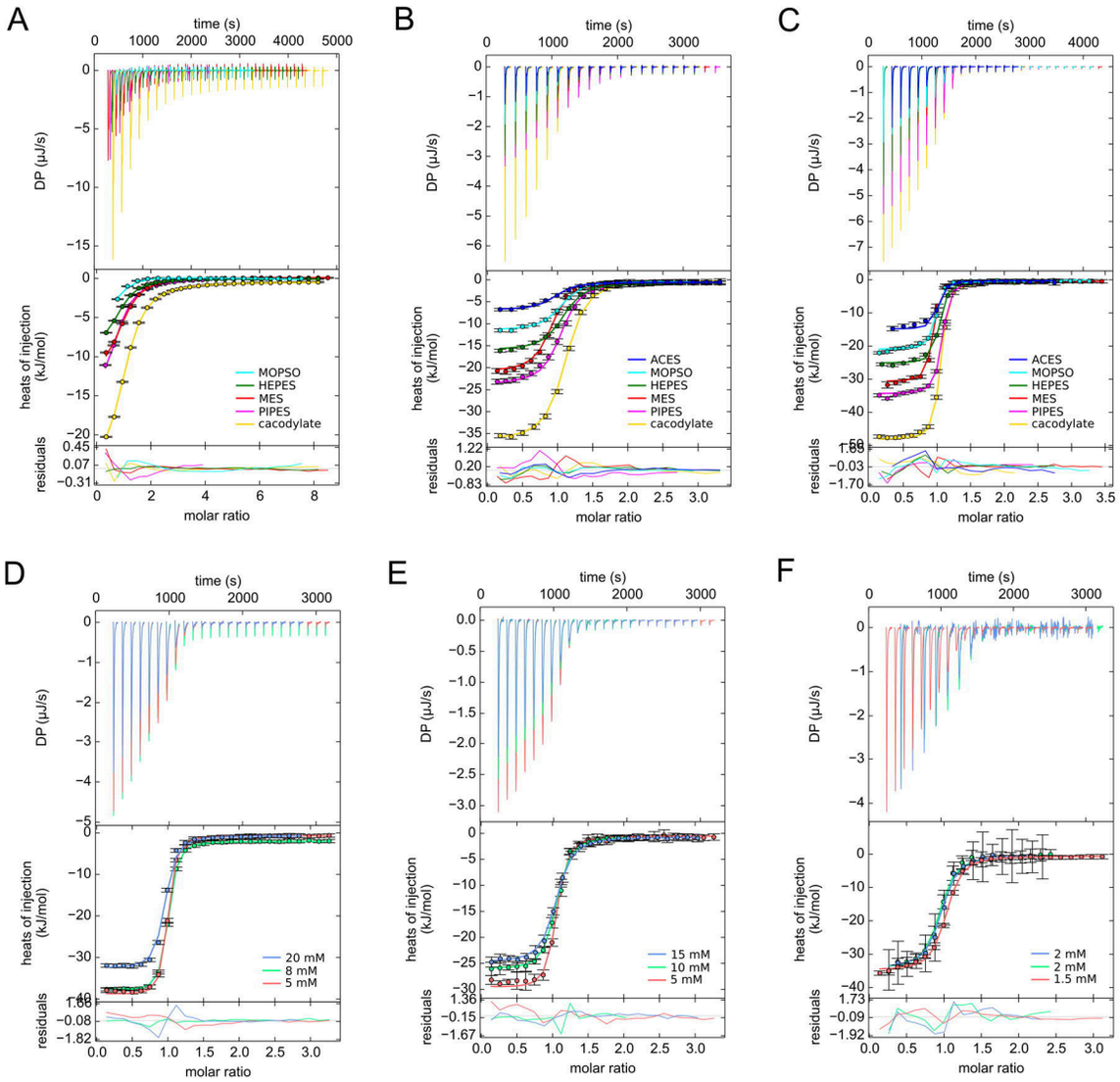


Figure S5.1. ITC thermograms of the **(A)** direct titration of **2**, **(B)** direct titration of **3**, **(C)** direct titration of **4**, **(D)** displacement titration of **1**, **(E)** displacement titration of **5**, **(F)** displacement titration of **6**. The direct titrations shown in panels A–C were performed in different buffers to correct $\Delta H°$ for superimposed buffer ionization reactions and to determine the proton transfer as shown in Figure S5.2. For the displacement titrations shown in panels D–F, the different isotherm colors indicate the different concentrations of weak ligand used for pre-incubation of TLN. For clarity reason always only one thermogram of the measurements performed at least in triplicate is shown. The thermodynamic binding profiles determined for the individual buffers are listed in Table S5.3 (**2**), Table S5.4 (**3**), and Table S5.5 (**4**). The results of the displacement titrations of **1**, **5**, and **6** are listed in Table S5.6.

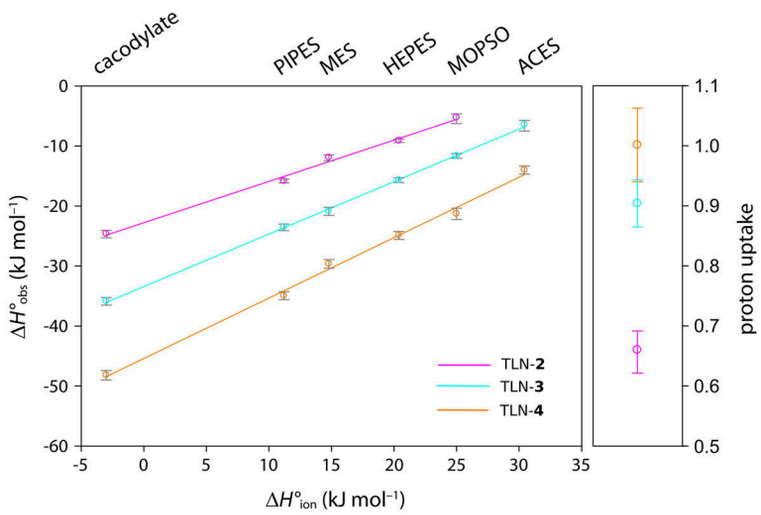


Figure S5.2. Determination of the number of protons transferred and the heat of buffer ionization upon complex formation of TLN-**2**, TLN-**3** and TLN-**4**. In the left diagram, the observed standard enthalpies upon ligand binding $\Delta H°_{obs}$ measured in six different buffers are plotted against the standard heat of ionization $\Delta H°_{ion}$ of the respective buffer [241]. Error bars in gray represent the 95.4% confidence interval of the $\Delta H°_{obs}$ values. Please note that the enthalpic signal of the complex formation of TLN-**2** was not measured in ACES buffer due to a too low exothermic signal. In the right diagram, the number of protons transferred upon complex formation of the three studied reactions is displayed as analyzed by global analysis with *SEDPHAT*. Error bars represent the 95.4% confidence interval. Data values are listed in Table S5.2.

Table S5.2. Proton transfer between buffer and TLN upon complex formation of TLN with ligands **4**, **3** and **2** as analyzed by global analysis with *SEDPHAT*.

Ligand	TLN proton uptake from buffer molecules (minimum and maximum of the 95.4% confidence interval)
4	1.001 (0.940, 1.063)
3	0.904 (0.865, 0.943)
2	0.660 (0.622, 0.692)

Table S5.3. Thermodynamic parameters obtained for **2**. All parameters are listed as rounded values.

	best-fit values global analysis (95.4% confidence interval)				locally fitted concentration correction factor[a]
buffer	K_d (μM)	$\Delta G°$(kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})	
buffer	40.750	-25.1	-22.7	-2.4	
corrected	(45.793, 36.299)	(-24.8, -25.3)	(-23.6, -21.9)	(-1.2, -3.4)	
cacodylate	39.604	-25.1	-24.7	-0.5	1.014, 0.997, 0.971
	(43.013, 36.425)	(-24.9, -25.3)	(-25.3, -24.1)	(0.4, -1.3)	
PIPES	52.301	-24.4	-15.8	-8.6	0.731, 0.754, 0.787
	(54.676, 50.038)	(-24.3, -24.5)	(-16.1, -15.5)	(-8.2, -9.0)	
MES	35.817	-25.4	-11.9	-13.4	0.927, 0.934, 0.951
	(41.314, 31.060)	(-25.0, -25.7)	(-12.5, -11.5)	(-12.5, -14.3)	
HEPES	39.526	-25.1	-9.1	-16.0	0.800, 0.797, 0.749,
	(43.301, 36.091)	(-24.9, -25.4)	(-9.4, -8.8)	(-15.5, -16.5)	0.796
MOPSO	16.239	-27.3	-5.3	-22.1	0.663, 0.561, 0.543
	(27.511, 9.335)	(-26.0, -28.7)	(-6.3, -4.7)	(-19.8, -24.0)	
ACES[b]	-	-	-	-	-

[a] The concentration correction factor is in this case equivalent to the *n* value. The number of active sites per molecule was set to 1, and the concentration of the ligand was considered as accurately known. Therefore, this value corrects for inaccurate protein concentrations, as well as for inactive protein fractions observed in some of the applied buffers. The number of listed concentration correction factors reflects the number of performed measurements. [b] The enthalpic signal in ACES buffer was too low for a reliable determination of the thermodynamic binding profile

Table S5.4. Thermodynamic parameters determined for **3**. All parameters are listed as rounded values.

	best-fit values global analysis (95.4% confidence interval; statistics minimum and maximum)				locally fitted concentration correction factor
buffer	K_d (μM)	$\Delta G°$(kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})	
buffer	0.992	-34.3	-33.7	-0.5	
corrected	(1.234, 0.792)	(-33.7, -34.8)	(-34.8, -32.7)	(1.1, -2.1)	
cacodylate	0.929	-34.4	-35.9	1.4	1.131; 1.065; 1.040
	(1.079, 0.798)	(-34.1, -34.8)	(-36.5, -35.2)	(2.5, 0.4)	
PIPES	1.358	-33.5	-23.5	-9.9	0.969; 1.005;0.992
	(1.599, 1.148)	(-33.1, -33.9)	(-24.1, -23.0)	(-9.0, -10.9)	
MES	0.710	-35.1	-20.9	-14.2	0.868; 0.835; 0.855
	(0.937, 0.530)	(-34.4, -35.8)	(-21.6, -20.2)	(-12.8, -15.6)	
HEPES	1.191	-33.8	-15.7	-18.1	1.005; 0.964; 0.956
	(1.447, 0.975)	(-33.3, -34.3)	(-16.1, -15.3)	(-17.2, -19.0)	
MOPSO	1.114	-34.0	-11.7	-22.3	0.991; 0.984; 0.940
	(1.427, 0.850)	(-33.4, -34.7)	(-12.0, -11.3)	(-21.3, -23.4)	
ACES	2.590	-31.9	-6.5	-25.4	0.939; 0.782; 0.742
	(5.013, 1.303)	(-30.3, -33.6)	(-7.5, -5.7)	(-22.7, -27.9)	

Table S5.5. Thermodynamic parameters determined for **4**. All parameters are listed as rounded values.

	best-fit values global analysis (95.4% confidence interval; statistics minimum and maximum)				locally fitted concentration correction factor
buffer	K_d (μM)	$\Delta G°$(kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})	
buffer	0.138	-39.1	-45.1	5.9	
corrected	(0.201, 0.090)	(-38.2, -40.2)	(-46.5, -43.7)	(8.3, 3.5)	
cacodylate	0.105	-39.8	-48.2	8.4	0.959; 0.939; 0.954;
	(0.141, 0.076)	(-39.1, -40.6)	(-49.0, -47.4)	(9.9, 6.8)	0.933
PIPES	0.124	-39.4	-34.9	-4.5	1.030; 0.954; 1.020;
	(0.168, 0.089)	(-38.7, -40.3)	(-35.6, -34.3)	(-3.0, -6.0)	0.989
MES	0.119	-39.5	-29.6	-9.9	0.895; 0.888; 0.871
	(0.174, 0.077)	(-38.6, -40.6)	(-30.4, -28.9)	(-8.2, -11.7)	
HEPES	0.182	-38.5	-24.9	-13.6	0.962; 0.961; 0.962
	(0.267, 0.118)	(-37.5, -39.5)	(-25.6, -24.2)	(-11.9, -15.3)	
MOPSO	0.207	-38.1	-21.3	-16.9	0.928, 0.924, 0.882
	(0.333, 0.122)	(-37.0, -39.5)	(-22.3, -20.4)	(-14.7, -19.1)	
ACES	0.279	-37.4	-14.0	-23.4	0.942; 0.892; 0.892
	(0.500, 0.132)	(-36.0, -39.3)	(-14.7, -13.3)	(-21.3, -25.9)	

Table S5.6. Thermodynamic data of **1**, **5** and **6** as determined by ITC displacement titrations in cacodylate buffer. All values are listed as rounded values.

	best-fit values global analysis (95.4% confidence interval; statistics minimum and maximum)				locally fitted concentration correction factor
ligand	K_d (mM)	$\Delta G°$(kJ mol^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})	
1[a]	5.659 (8.321, 3.762)	-12.8 (-11.9, -13.8)	-	-	0.907, 0.944, 0.950
5[a]	2.407 (3.652, 1.507)	-14.9 (-13.9, -16.1)	-	-	1.000, 0.998, 0.981
6[b]	0.336 (0.752, 0.168)	-19.8 (-17.8, -21.5)	-16.9 (-22.7, -13.5)	-2.9 (4.9, -8.0)	0.981, 0.894, 0.916

[a] Thermodynamic parameters $\Delta H°$ and $-T\Delta S°$ could not be determined accurately due to the extremely low affinity of **1** and **5**, and are thus omitted. [b] Thermodynamic parameters of **6** were not corrected for the heats of buffer ionization.

5.9.2 Analysis of the experimentally phased TLN-1

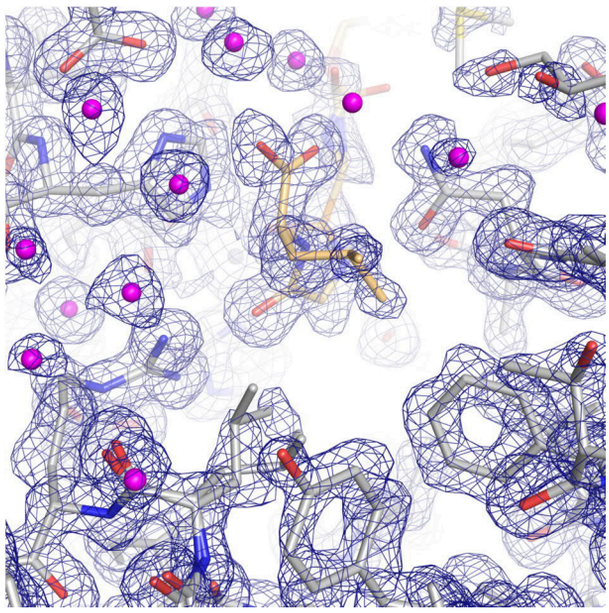


Figure S5.3. Experimentally phased absolute-scale electron density map of TLN-**1** shown as blue mesh at a contour level of 0.7 $e^-/Å^3$. For orientation, the refined model of TLN-**1** is superimposed onto the electron density map. TLN residues are displayed as gray sticks, the bound ligand is displayed as light orange stick model and water molecules are displayed as magenta spheres.

Table S5.7 (related to Figure 5.2). Solvent excluded volume and electron content of selected cavities of TLN-**1**.

cavity	volume (Å^3)	integrated electron content (e^-)[a]	water molecules in the model
S_1'	141	6.6	0
reference cavity 1	59	19.4	3
reference cavity 2	93	13.4	1
reference cavity 3	16	0.9	0

[a] All integrated electron contents were scaled on the intrinsically bound zinc atom of TLN.

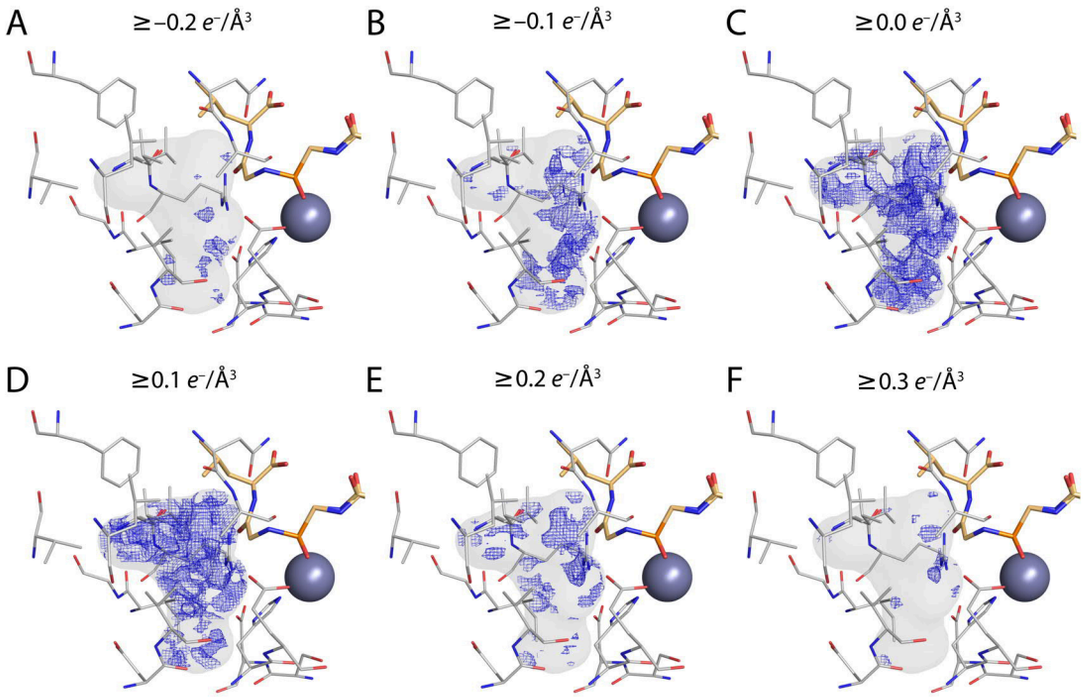


Figure S5.4 (related to Figure 5.2). Absolute-scale electron density map (blue mesh) within the S_1' cavity at different contour levels as labeled in panels A–F.

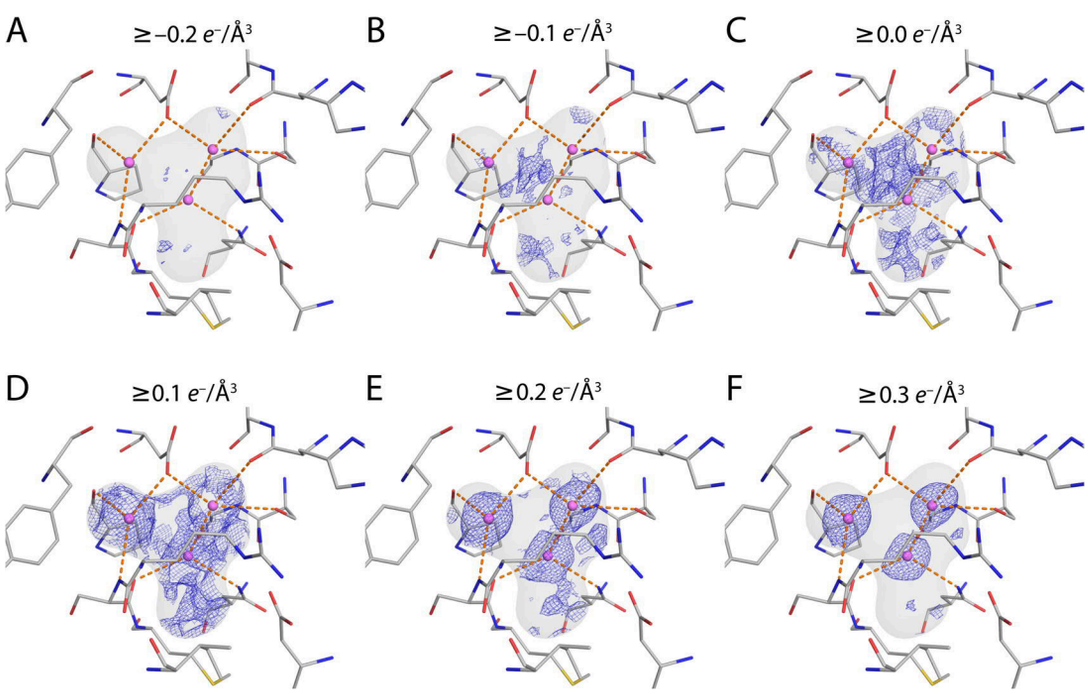


Figure S5.5 (related to Figure 5.2). Absolute-scale electron density map (blue mesh) within the reference cavity 1 at different contour levels as labeled in panels A–F.

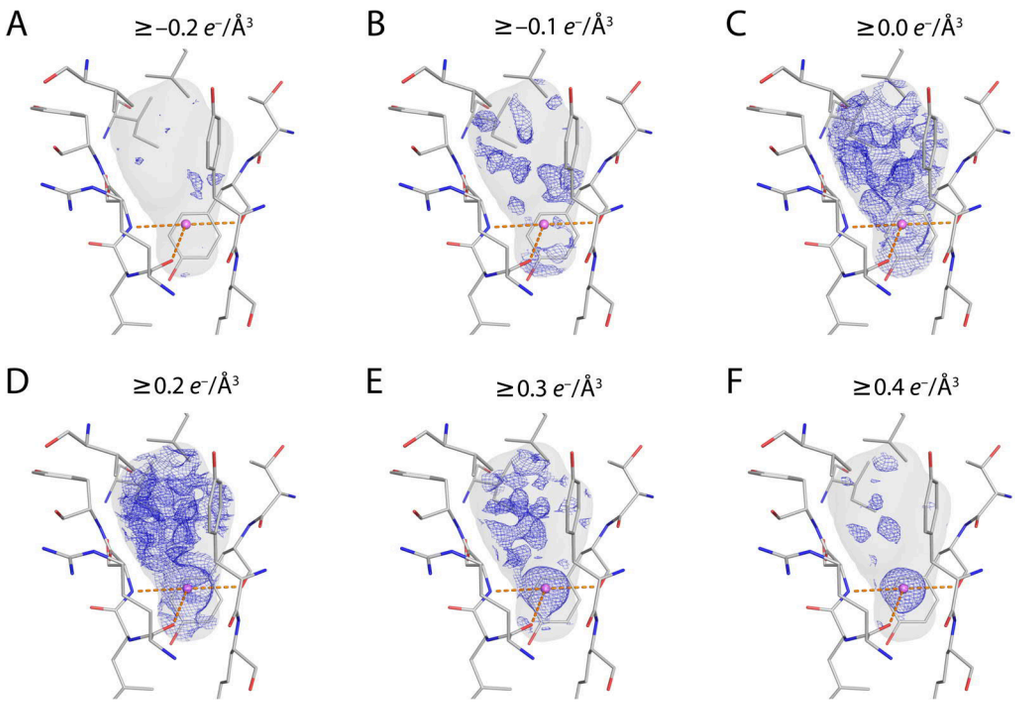


Figure S5.6 (related to Figure 5.2). Absolute-scale electron density map (blue mesh) within the reference cavity 2 at different contour levels as labeled in panels A-F.

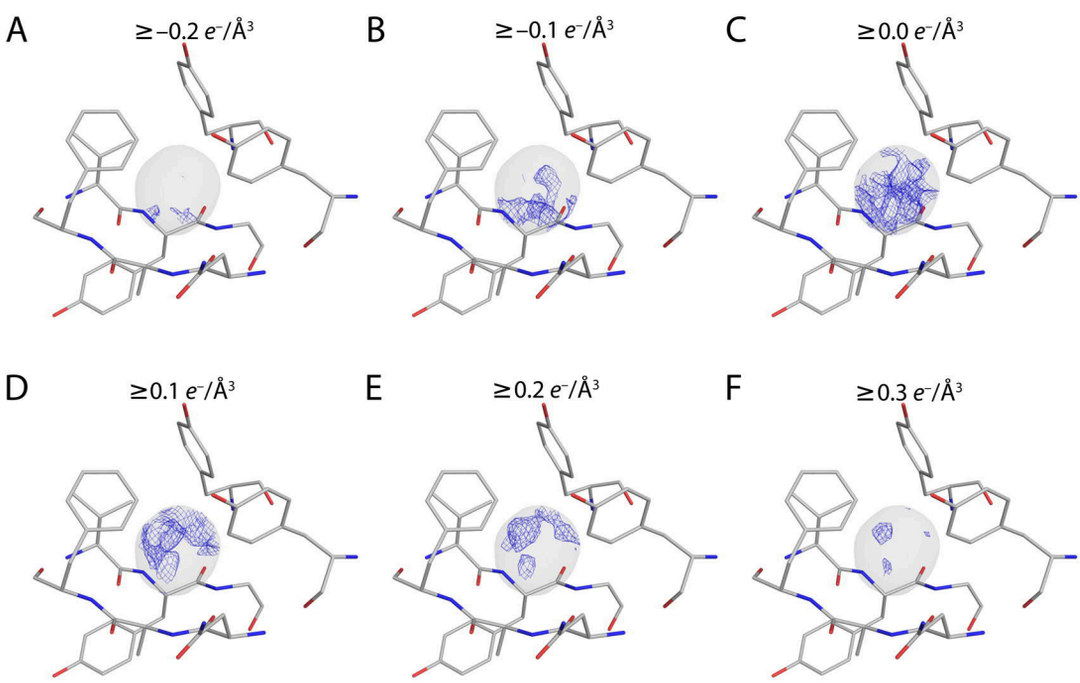


Figure S5.7 (related to Figure 5.2). Absolute-scale electron density map (blue mesh) within the reference cavity 3 at different contour levels as labeled in panels A-F.

5.9.3 **Crystal structure of TLN-1-Kr**

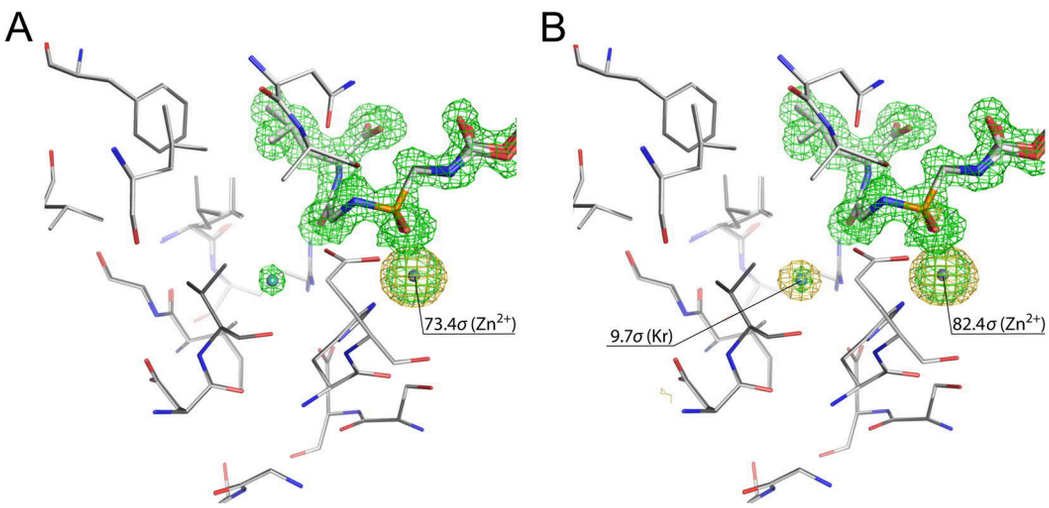


Figure S5.8. Krypton binding site in the S_1' cavity of the crystal structure TLN-**1**-Kr. **(A)** Dataset collected below the krypton absorption edge. **(B)** Dataset collected above the krypton absorption edge. In both panels, the crystal structure of TLN-**1**-Kr (dark gray) is superimposed on the crystal structure of the native TLN-**1** (light gray). The anomalous map is shown in gold at a contour level of 5σ, and the detected peaks are labeled with their intensity maximum. The F_o–F_c omit map of the krypton atom, zinc ion and the ligand is shown in green at a contour level of 3σ

.

Chapter 6

Elucidating the Origin of Long Residence Time Binding for Inhibitors of the Metalloprotease Thermolysin

Jonathan Cramer,[1,*] Stefan G. Krimmer,[1,*] Veronica Fridh,[1,*] Tobias Wulsdorf,[1] Robert Karlsson,[2] Andreas Heine,[1] Gerhard Klebe[1]

*These authors contributed equally

[1]Department of Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 6, 35032 Marburg, Germany

[2]GE Healthcare Bio-Sciences AB, SE-751 84 Uppsala, Sweden

ACS Chemical Biology **2017**, 12: 225–233

6.1 Abstract

Kinetic parameters of protein–ligand interactions are progressively acknowledged as valuable information for rational drug discovery. However, a targeted optimization of binding kinetics is not easy to achieve and further systematic studies are necessary to increase the understanding about molecular mechanisms involved. We determined association and dissociation rate constants for 17 inhibitors of the metalloprotease thermolysin by surface plasmon resonance spectroscopy and correlated kinetic data with high-resolution crystal structures in complex with the protein. From the structure–kinetics relationship, we conclude that the strength of interaction with Asn112 correlates with the rate-limiting step of dissociation. This residue is located at the beginning of a β-strand motif that lines the binding cleft and is commonly believed to align a substrate for catalysis. A reduced mobility of the Asn112 sidechain owing to an enhanced engagement in charge-assisted hydrogen bonds prevents the conformational adjustment associated with ligand release and transformation of the enzyme to its open state. This hypothesis is supported by kinetic data of ZF^{P}LA, a known pseudopeptidic inhibitor of thermolysin, which blocks the conformational transition of Asn112. Interference with this retrograde induced-fit mechanism results in variation of the residence time of thermolysin inhibitors by a factor of 74,000. The high conservation of this structural motif within the M4 and M13 metalloprotease families underpins the importance of this feature and has significant implications for drug discovery.

6.2 Introduction

Ultimately, the goal of a drug design project is the development of a clinical candidate that is efficacious *in vivo*. The main surrogate parameter for efficacy in the early stages of development is the affinity of a prospective lead candidate, assessed in an assay working under a thermodynamic equilibrium regime. In many cases, however, equilibrium binding affinity translates poorly to *in vivo* activity. Thus additional parameters such as the binding kinetics of protein–ligand interactions can add further insights and are increasingly appreciated as a valuable indicator for therapeutic efficacy [1,115,242]. Recent studies have shown that an increased receptor residence time ($\tau = 1/k_d$) translates into an enhanced pharmacological effect also in an *in vivo* setting [24,27]. Furthermore, in a study of adenosine A_{2A} receptor antagonists, drug response was more strongly correlated to variations in residence time than affinity under equilibrium conditions (K_D) [243]. Yet, the intentional modification of the kinetic parameters of a given ligand is not easy to achieve, because the underlying molecular determinants are poorly understood and thus a reliable structure–kinetics relationship is difficult to establish. Only in rare cases it is possible to rationalize binding kinetic parameters with respect to variations in the structure of an inhibitor in a mechanistically coherent way [150,244,245]. Generally, high molecular weight, pronounced lipophilicity (clogP >5) and enhanced molecular flexibility (number of rotatable bonds >5) are accepted to cause slow dissociation rates [246]. However, since these parameters are likely to correlate with unfavorable pharmacokinetic properties, they are not necessarily the most preferred ones to be considered and improved in a targeted drug optimization process. Computational analyses of buried water clusters in the binding pocket of the adenosine A_{2A} receptor in complex with a series of antagonists have shown that the number and position of thermodynamically unfavorable water molecules correlates with decreasing residence time of the inhibitor [247]. Computational analyses have suggested that desolvation or resolvation of the binding site prior to or after the binding event determine the rate-limiting steps for association or dissociation, respectively [174,248]. Especially the hydrophobic shielding of buried polar atoms in the protein binding site seems to prevent rapid rehydration and thus prolongs dissociation of a bound ligand [249].

Fluctuations in protein conformation can impose a major influence on the binding kinetic profile of an inhibitor [20,250]. For enoyl-ACP reductase it has been demonstrated that upon binding of the inhibitor PT70 a loop region that is disordered in the uncomplexed state transforms into an ordered helical structure. This reduces the association and dissociation rates of the binding event significantly [251]. The kinetic selectivity of the marketed drug thiotropium toward the M3 isoform of the muscarinic G protein coupled receptor is believed to originate from differences in the dynamic behavior of the ECL2 region relative to the M2 isoform [252]. Another common motif is the occlusion of the ligand binding site by a

hydrophobic lid [253,254]. The ability of a ligand to stabilize this closed conformation is reflected in its kinetic profile.

In order to further trace putative mechanisms responsible for extended residence times on the molecular level, we used the metalloprotease thermolysin (TLN) from *B. thermoproteolyticus* as a model system. This enzyme has been successfully consulted in the past as a target surrogate in the development of inhibitors for angiotensin converting enzyme and neutral endopeptidase [35–37]. Its chemical robustness, excellent crystallographic properties and ready access in large quantities renders TLN an ideal test system to study details of protein–ligand interactions.

The conformational dynamics of TLN and other zinc proteases upon ligand binding have been subject of several studies [160,255,256]. It has been hypothesized that TLN, analogous to several other members of the bacterial neutral protease family, undergoes a kind of "hinge-bending" motion about its central α-helix and folds upon the bound substrate. However, experimental evidence for this hypothesis remained elusive for a long time. Approaches that rely on a static crystallographic model have notoriously been hampered by the fact that a dipeptide (Val-Lys), cleaved off from the C-terminus of the protease by autoproteolysis, blocks access to the active site and stabilizes the closed conformation [160]. Exchange of the central zinc ion against other more heavy metal ions finally revealed an alternate conformation that was later confirmed as the putative *apo* form of TLN [255,256]. Although the overall difference between the two conformers is small (rmsd = 0.62 Å based on 316 C_α atoms) [255], larger conformational changes were observed for the side chains of Met120, Glu143, and Leu144.

In this study, the kinetic binding parameters of 17 closely related TLN inhibitors are analyzed with respect to slight chemical modifications in their partly solvent–exposed P_2' substituents (Figure 6.1A). The ligand series consists of three subsets that can be distinguished according to their C-terminal functionality. Either a charged carboxy group (subset *a*, blue), a neutral carboxamide (subset *b*, green) or an apolar aliphatic substituent (subset *c*, red) were considered. The kinetic profiles are correlated with high-resolution crystal structures of the compounds in complex with TLN (mean resolution across all 17 structures: 1.30±0.15 Å).

6.3 Results

6.3.1 SPR data collection

The kinetic binding parameters of **1–17** were determined by SPR using an assay based on capture of biotinylated TLN as outlined in the experimental section. The sensorgrams

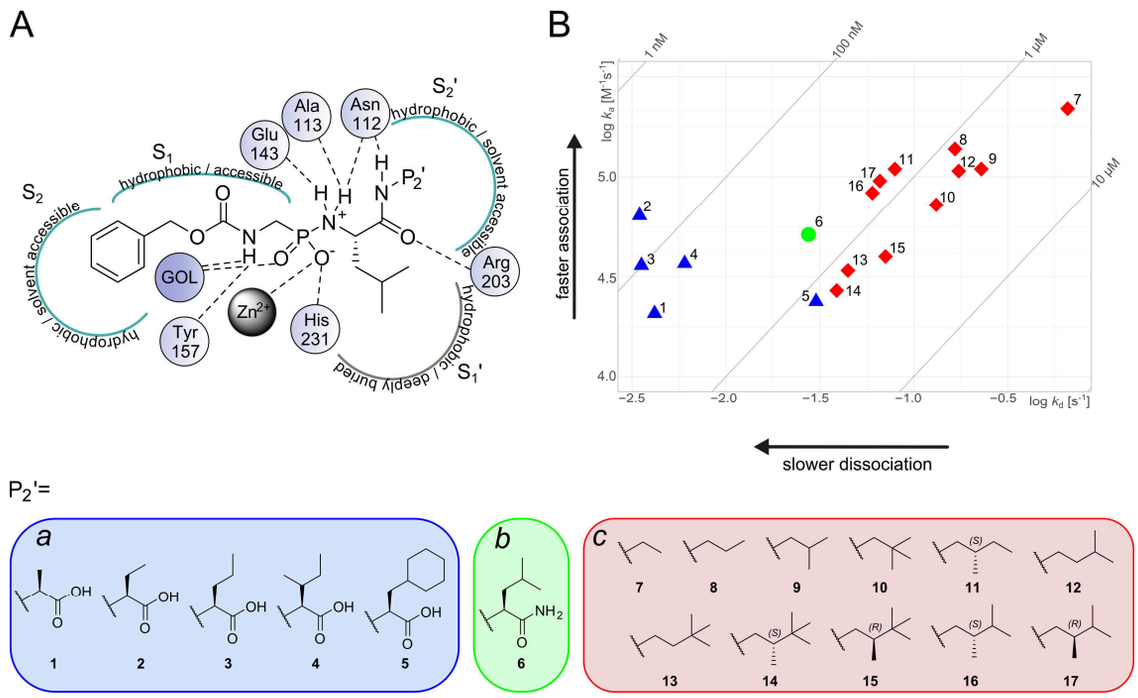


Figure 6.1. Binding kinetic data. **(A)** Schematic representation of the relevant interactions between phosphonopeptide ligands **1**–**17** and TLN. The ligands are categorized in three groups according to the chemical composition of their P_2′ groups. Subset *a* (blue): P_2′-carboxy (**1**–**5**), subset *b* (green): P_2′-carboxamide (**6**), subset *c* (red): P_2′-aliphatic (**7**–**17**). **(B)** Kinetic map (log k_a vs. log k_d) of **1**–**17**. Subset *a*: blue triangles, subset *b*: green circle, subset *c*: red diamonds. The diagonal lines indicate regions of equal affinity.

showed clear differences in binding characteristics within and/or between the three investigated series (Supporting Information Figure S6.1).

6.3.2 Kinetic data analysis

Figure 6.1B shows a map of the binding kinetic parameters (log k_a vs. log k_d) of ligands **1**–**17**. The diagonal lines indicate lines of equal affinity (equilibrium dissociation constant $K_D = k_d/k_a$). Overall, the affinity varies within the ligand series from 40 nM to 3.3 µM. The association rate constants across all subsets scatter within one order of magnitude about a mean value of 10^4 $M^{-1}s^{-1}$. In contrast, the dissociation rate constants display a larger variance between 6.15×10^{-1} s^{-1} and 2.77×10^{-3} s^{-1}. It is apparent that ligands bearing a C-terminal carboxy group (**1**–**5**) show a slower dissociation rate than ligands with only an aliphatic substituent at the P_2' position lacking the acid group (**7**–**17**). Compound **6** with a terminal carboxamide falls in-between both series with a k_d of 4.46×10^{-2} s^{-1}. Within subset *a*, **5** shows a significantly faster dissociation from the TLN binding pocket, with k_d increased by a factor of ten relative to the other ligands of the subset. In subset *c*, **13** and **14** depart toward slower dissociation and **7** displays the fastest

association and dissociation rates compared to the other compounds. The rapid approach to steady state for **7**, resulting in a squared shaped pulse (Supporting Information Figure S6.1), is governed mostly by its fast dissociation rate ($k_d = 6.15\times10^{-1}$ s^{-1}). A small subset of the inhibitors (**14**–**17**) was already analyzed previously in a detailed study including structural, thermodynamic and kinetic properties [16]. While differences in k_a were observed, we mainly focused on an interpretation of dissociation kinetics for this study.

In order to assess the dependence of the binding kinetic data of the phosphonamidates on electrostatic properties, the rate constants of a representative member of each subset were determined at three different ionic strength conditions (Supporting Information Table S6.2). In addition to the standard buffer conditions, NaCl was added in a concentration of 100 mM or 500 mM, respectively. A significant deviation could not be detected in the investigated range of ionic strength.

6.3.3 **X-ray crystallography**

The high-resolution crystal structures of several ligands in complex with TLN have been reported by us in previous studies [15,93,94]. The structures of **5** and **6** were additionally determined in the context of this study (crystallographic information in Supporting Information Table S6.3). In Figure 6.2A, **1** is depicted as a representative example for the binding mode topology of **1**–**17** within the binding cleft of TLN. The carbamate nitrogen interacts with Tyr157 and a glycerol (GOL) molecule that is picked up from the buffer and bound to the S_2 pocket. One oxygen atom of the phosphonamidate moiety binds to the catalytic zinc ion in a monodentate fashion and is further positioned within hydrogen-bond distance of His231. The other oxygen atom interacts with the GOL molecule. The phosphonamidate nitrogen is likely protonated under the applied experimental conditions [135] and interacts with Glu143, Ala113 and Asn112. The hydrophobic leucine side chain of the inhibitors is deeply buried in the S_1' pocket. Additionally, a hydrogen-bonding interaction of Arg203 to the leucine carbonyl oxygen is established. The polar groups of the C-terminal ligand portion are in contact with the side chain carboxamide of Asn112 (*vide infra*). The variable hydrophobic P_2' moieties are positioned in the shallow, solvent exposed S_2' pocket. Figure 6.2B shows a superposition of the binding mode of all investigated ligands. Apart from the chemically deviating P_2' substituents, the structures show virtually perfect overlap of the parent inhibitor scaffold.

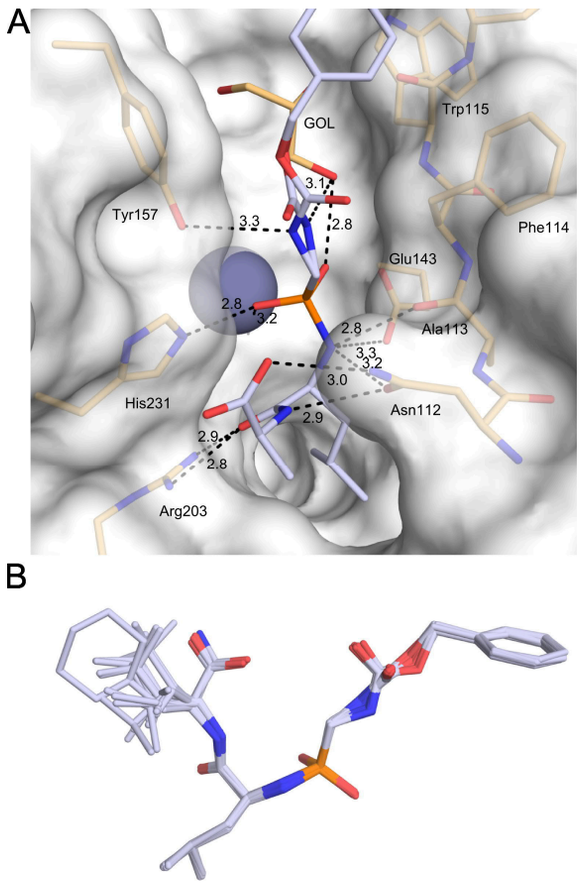


Figure 6.2. General binding mode of phosphonopeptide ligands to TLN. **(A) 1** is shown in blue, protein residues adjacent to the inhibitor are shown in orange. The dashed lines indicate hydrogen bonds between ligand and protein with the distance between the respective heavy atoms annotated in Å. The gray-blue sphere represents the catalytic zinc ion. The solvent-excluded surface of TLN is shown in white. **(B)** Superposition of the binding mode of inhibitors **1**–**17**.

6.4 Discussion

6.4.1 Interactions of inhibitors 1–17 with Asn112

Since all interactions that do not involve the P_2' substituent are geometrically highly conserved between **1**–**17**, the explanation for their deviating k_d values must originate from the interactions of their deviating C-terminal portion. A comparison of the interactions across the three subsets shows that the ligands engage in a deviating hydrogen-bonding pattern with respect to the side chain of Asn112. As shown in Figure 6.3, all inhibitors involve the carboxamide group of this residue into multiple hydrogen bonds. In all cases, the phosphonamidate nitrogen, which is likely protonated under the applied conditions,

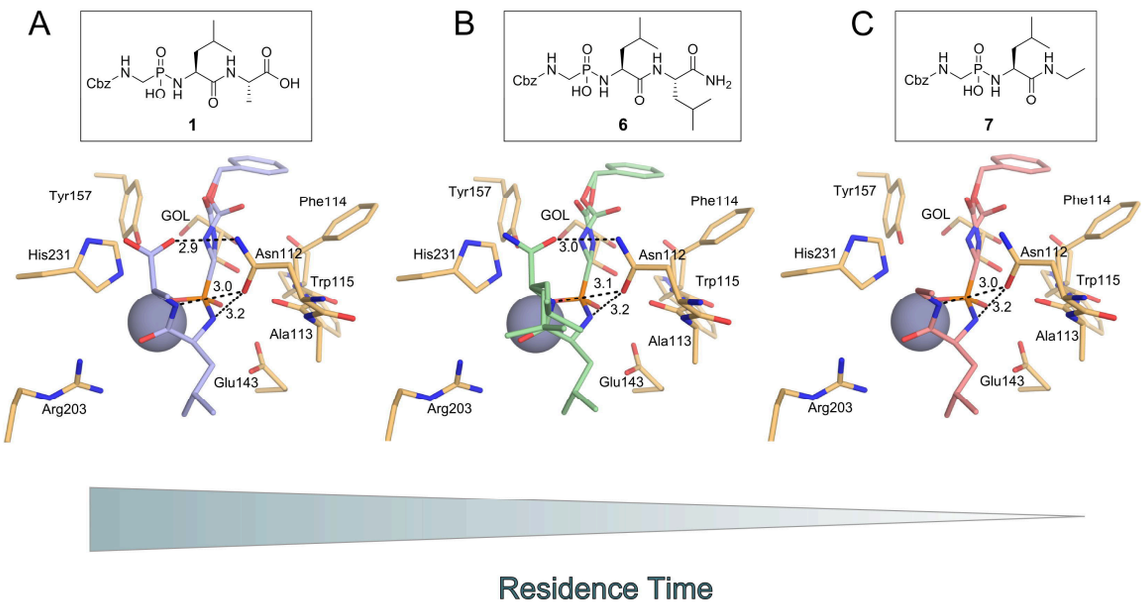


Figure 6.3. Interaction pattern of a representative ligand of each group with Asn112. Residues of the active site are shown in orange. For reasons of clarity, only hydrogen bonds formed by Asn112 are shown. **(A)** The C-terminal amino acid moiety of P_2'-carboxy inhibitor **1** (blue) engages in a bidentate charge-assisted hydrogen bond to Asn112. **(B)** The general binding mode of P_2'-amide inhibitor **6** (green) resembles that of P_2'-carboxy compounds. **(C)** P_2'-aliphatic inhibitor **7** (red) solely interacts with the carbonyl oxygen of Asn112. Bidentate coordination of the C-terminal residue is lost.

interacts with the side chain carbonyl oxygen of Asn112. The secondary amide nitrogen of the P_2' group also forms a hydrogen bond to Asn112. Inhibitors with an aliphatic P_2' group do not show any further interactions with Asn112 (Figure 6.3C). The addition of a C-terminal carboxy group allows further fixation of Asn112 due to a chelating charged-assisted hydrogen bond to the side chain carboxamide nitrogen (Figure 6.3A). The binding mode of **6**, which features a C-terminal primary carboxamide, resembles that of the analogous carboxy compounds (Figure 6.3B). In this case, however, the hydrogen bond between the terminal carboxamide and Asn112 is weakened due to the loss of a formal charge, which potentially induces an electrostatic enhancement in case of **1–5**.

6.4.2 The conformational dynamics of Asn112 and their implications for the kinetics of TLN

Previous discussions concerning the dynamic properties of TLN mainly focused on conformational transformations of Met120, Glu143 and Leu144 [255,256]. So far, a contribution of Asn112 has not been taken into consideration. Our structural and kinetic results, however, suggest major involvement of this residue in the dissociation step. We

therefore propose the following release mechanism: The rotation of Asn112 is necessary to enable the "hinge-bending" motion, which is required to expel a tightly bound substrate or ligand from the binding pocket. In Figure 6.4A, the structure of TLN in complex with the peptidic inhibitor IMPI (insect metalloprotease inhibitor) from *Galleria mellonella* is shown. IMPI is cleaved by the protease, yet, the hydrolysis product is not released but forms an inhibitory complex with the enzyme [257]. Therefore, its binding mode supposedly resembles a geometry adopted by the substrate after peptide-bond cleavage. The C-terminal

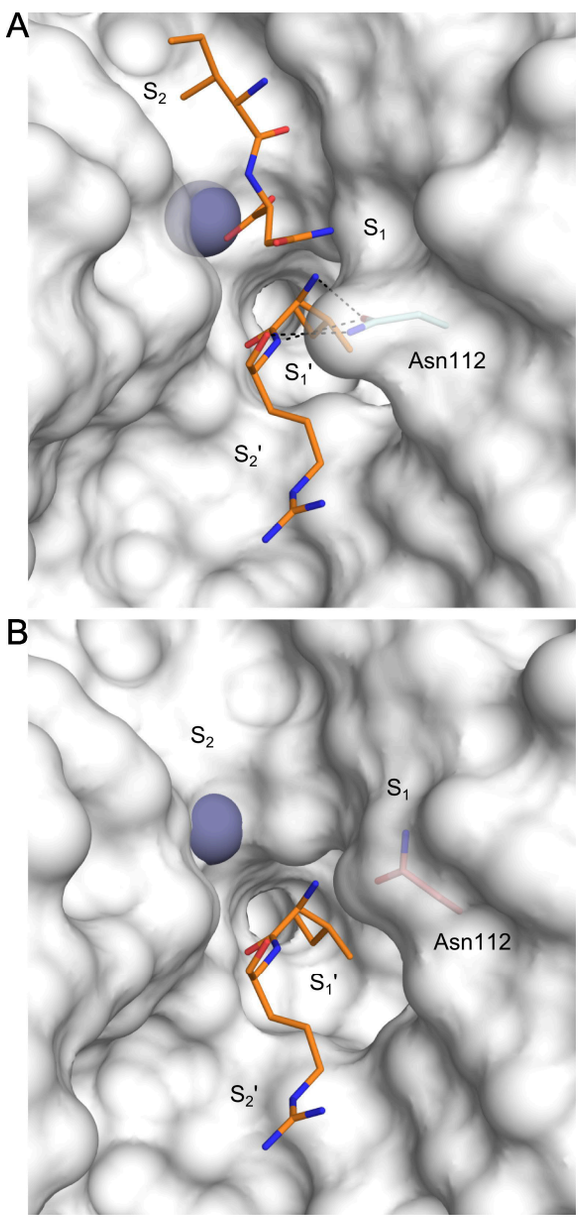


Figure 6.4. Binding mode of the cleaved peptidic product fragments of IMPI (pdb entry: 3ssb). **(A)** Only the four central residues of the "reactive-site" loop are shown in orange. Asn112 in the closed conformation is shown in cyan. **(B)** Superposition of IMPI residues 57–58 with the open conformation of TLN. Asn112 is shown in red.

fragment of IMPI engages in a contact to Asn112 resembling the hydrogen-bonding pattern also observed for our P_2'-carboxy ligands. Obviously, this interaction stabilizes Asn112 in a position, in which its carboxamide side chain forms a lid over the S_1' pocket, augmenting the barrier for the release of the bound substrate product. If the latter complex is superimposed with the open conformation of TLN (Figure 6.4B), it becomes apparent that the outward rotation of Asn112 is involved in an opening of the S_1' specificity pocket, which subsequently allows the release of the cleaved peptide fragment. Our hypothesis supports the assumption that the dissociation of a cleaved substrate follows a sequential process. As a first step, the weakly bound N-terminal peptide (occupying the less specific binding pockets on the unprimed side) has to dissociate from the binding pocket. Once the S_1 pocket is abandoned (and conceivably rehydrated), Asn112 is able to rotate toward this pocket, thus giving access to the well-defined S_1' pocket. This retrograde induced-fit mechanism allows the dissociation of the deeply buried C-terminal product fragment from the hydrophobic cavity. In contrast to this sequential process, a transition-state mimetic inhibitor would have to dissociate from the binding pockets in a concerted way. A pronounced fixation of Asn112 through a strong interaction to the P_2' substituent imposes an additional energetic barrier for the induced-fit step along the escape trajectory [20].

A structural alignment of proteases from the M4 family (TLN-like zinc metalloproteases) reveals that Asn112 is highly conserved across the members of this group (97% consensus in Pfam-A family) [258]. Within the less related M13 family (Neprilysin family), multiple proteases featuring this structural motif are found (79% consensus in Pfam-A family). Figure 6.5 shows the crucial part of the secondary structure alignment of TLN with other members of the M4 and M13 family. Important representatives such as the human endothelin-converting-enzyme (ECE-1) or the human neutral endopeptidase (Neprilysin), associated

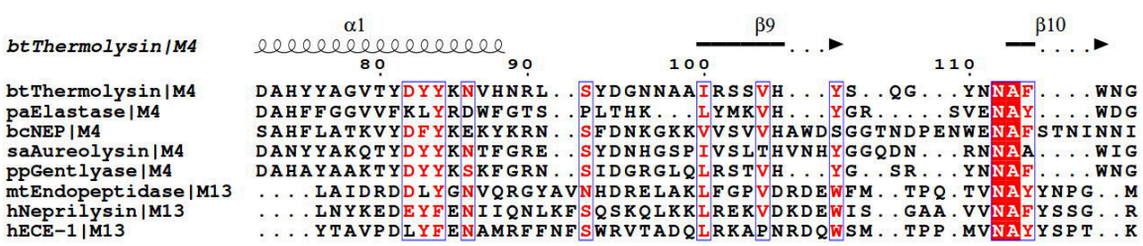


Figure 6.5. A section of the secondary structure alignment of TLN and representative members of the M4 and M13 protease families (alignment of the full sequence in SI). Partially or functionally conserved residues are indicated by red letters, full conservation is highlighted with white letters on red background. Lower case letters indicate the source organism of the respective protein (bt = *B. thermoproteolyticus*, pa = *P. aeruginosa*, bc = *B. cereus*, sa = *S. aureus*, pp = *P. polymyxa*, mt = *M. tuberculosis*, h = human). For the secondary structure alignment the protein structure comparison service PDBeFold at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm) was used [298]. The results were displayed using the ESPript server (http://espript.ibcp.fr) [299].

with diseases like breast cancer, Alzheimer's disease and heart failure belong to these families [259–261]. Furthermore, it has been shown that N112X mutants (X = A, H, K, R) of TLN lose their catalytic activity [262]. Only the variants N112D and N112E partly retain their proteolytic activity (64% and 19% respectively) [263]. This underpins the crucial role of Asn112 for the substrate turnover of TLN-like proteases.

6.4.3 **Structure-Kinetics relationship**

Based on these considerations, it becomes obvious that the strength of the ligand-to-Asn112 interaction is reflected by the kinetic binding profile. Especially the dissociation rate constant correlates with this property. Members of subset *a* engage in strong, charged-assisted hydrogen bonds between the ligand carboxy and Asn112 carboxamide group. Consequently, these inhibitors show the longest residence times across the series. For **6**, the chelating hydrogen bond is attenuated due to the loss of one formal negative charge. As a result, its k_d corresponds to a three-fold faster release (compared to the mean of **1**–**5**). Total abandonment of this interaction, as given for ligands from subset *c*, results in a thirteen-fold accelerated dissociation (factor between the mean k_d value of **1**–**5** and **7**–**17**).

In a study by Bartlett and Marlowe [264], the binding kinetics of a related inhibitor exhibiting a benzyl group in the P_1 position (ZF^{P}LA, **18**, Figure 6.6) was determined in a photometric inhibition assay. The latter ligand is the most potent TLN inhibitor described in literature, with an affinity of K_D = 68 pM. This high potency originates from a considerably lower dissociation rate resulting in the remarkable residence time of 168 days ($k_d = 6.8\times10^{-8}$ s^{-1}). Although inherent differences between the applied photometric method and our SPR analysis to record the kinetic constants hampers a straightforward quantitative comparison, the data of **1**, examined in both studies, suggests that the deviations between both approaches fall maximally into the range of one order of magnitude (for a detailed comparison see Supporting Information Figure S6.2). The exchange of the substitution pattern at P_1 from hydrogen to (*R*)-benzyl resulted in a 74,000-fold deceleration of the dissociation rate constant, based on the photometric inhibition assay (Figure 6.6). This finding is in full agreement with our proposed release mechanism and explains the remarkable kinetic properties of **18**. The binding mode shows some significant deviations from **1**–**17** that all contain a Gly^P-motif in the P_1 position. In the crystal structure of **18**, the Phe^P substituent is positioned in van der Waals distance to Phe114 (Figure 6.7A). The mutual positioning of the two aromatic portions blocks the outward rotation of Asn112. However, as described, this movement is an important prerequisite for the simultaneous release of the bound ligand, while translocating TLN to its open conformation. This hypothesis is substantiated by a structural superposition of the open conformation of the *apo*-protein and the inhibited complex of **18** (Figure 6.7B).

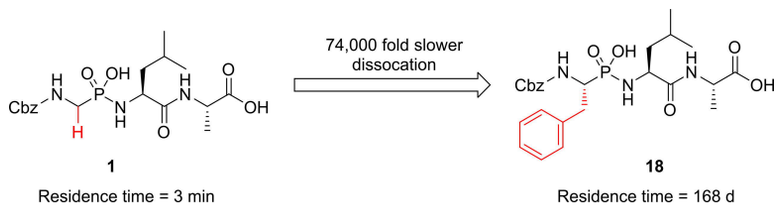


Figure 6.6. Influence of a P_1 substitution from hydrogen to *(R)*-benzyl on residence time (residence time for **18** was calculated from k_d determined by Bartlett and Marlowe) [264].

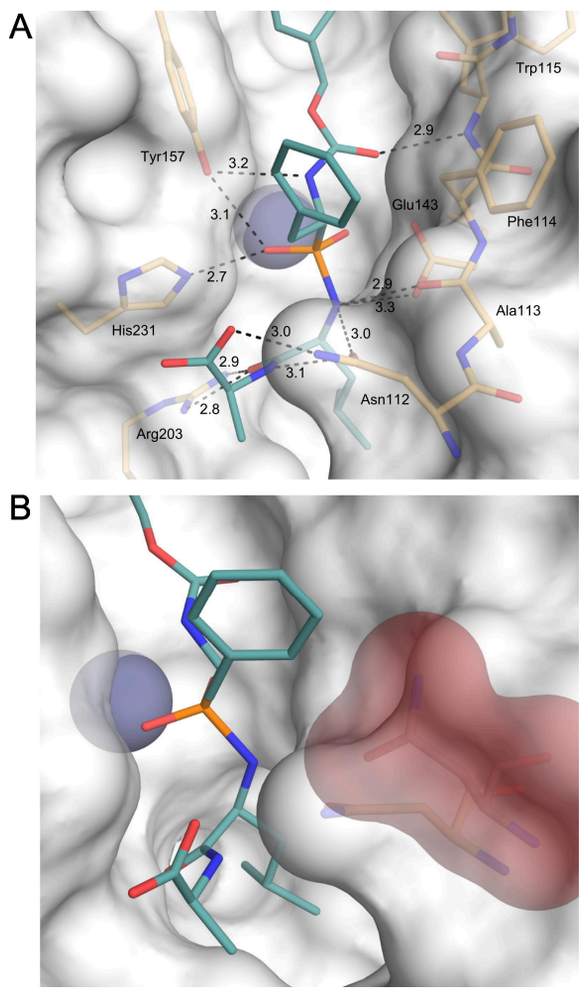


Figure 6.7. Binding mode of ZF^{P}LA (**18**; PDB entry 4TMN). **(A)** The Phe^P substituent is within van der Waals distance to Phe114 and shields the top of the binding cleft. **(B)** Superposition of Asn112 from TLN in the open conformation (red). In the closed conformation, the polar side chain of Asn112 is in close contact to the apolar Phe^P portion of ligand **18**.

Within each subset, binding kinetic parameters are fine-modulated and a more complex picture determines the correlation. As discussed in our previous contribution, the quality and completeness of the surface water network wrapping around ligand-exposed parts of the newly formed protein–ligand complex can take impact on k_d in the range of one order of magnitude [16]. Within subset *a*, **5** is clearly separated from **1**–**4** in the direction of faster dissociation. Since **5** exhibits the largest molecular weight within the series, this result opposes the generally accepted correlation of increasing molecular mass with slower dissociation rate [246]. A closer inspection of the binding modes of **1**–**5** reveals differences in the surface water network adjacent to the partly exposed P_2' side chain of the ligands, which provide an explanation (Figure 6.8). Due to the large size of the hydrophobic cyclohexyl substituent of **5**, the formation of the surface water network remains incomplete and is more strongly perturbed compared to those formed with **1**–**4**. This destabilization of the surface water network supposedly facilitates

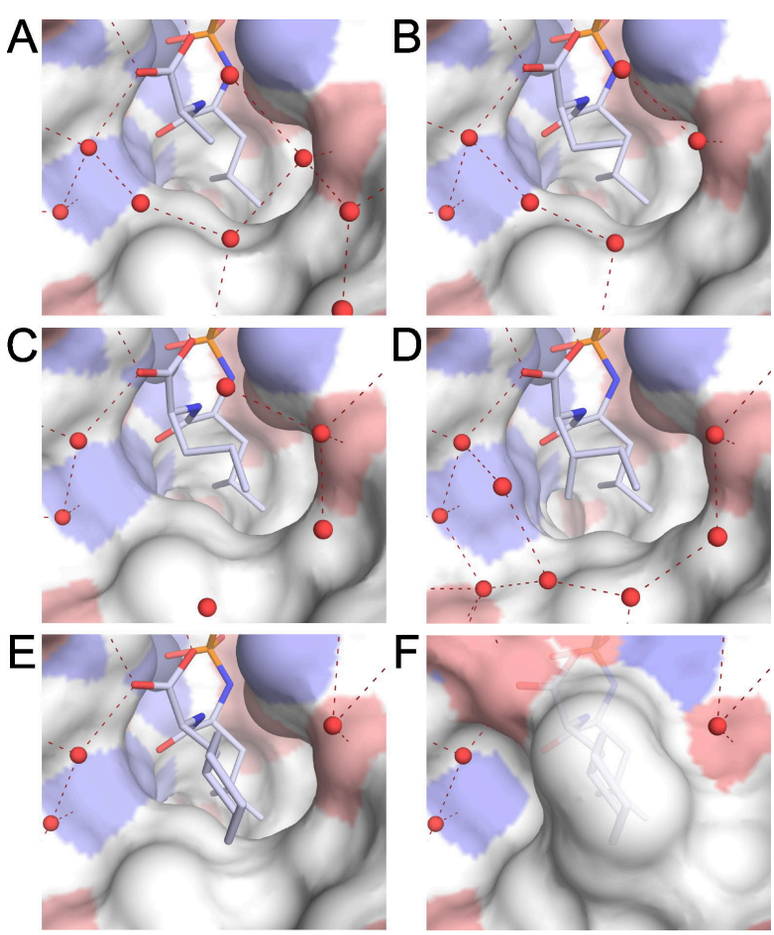


Figure 6.8. Water arrangement around the P_2'-portion of ligands. **(A) 1**, **(B) 2**, **(C) 3**, **(D) 4** and **(E) 5** in complex with TLN. **(F)** Representation of the solvent-excluded surface generated from the complex of **5** with TLN. The surface is color-coded according to atom types (red = oxygen, blue = nitrogen, white = carbon). Water molecules are shown as red spheres with red dashes indicating hydrogen bond interactions.

dissociation of the bound ligand due to the loss of solvent–solvent interactions that have to be disrupted upon ligand dissociation.

It has been demonstrated for several protein–ligand systems that electrostatic attraction or repulsion can play an important role in binding kinetics [265–268]. While electrostatic factors are generally believed to predominantly influence association through enhanced probability for diffusional collisions or conformational preorganization of the binding partners, minor impact on the dissociation process cannot be neglected. In order to elucidate possible long-range electrostatic effects between the differently charged molecules, we repeated our kinetic experiments in buffers of varying ionic strength. In our experiments, none of the kinetic constants for the TLN-phosphonamidate interactions showed a significant dependence on the ionic strength, regardless of the presence or absence of any nominal charge on the ligand's C-terminus (Supporting Information Table S6.2). We are therefore confident that long-range electrostatic effects do not matter significantly in the current example and the differences in binding kinetics can be traced back to the interaction with Asn112.

6.5 Conclusion

The kinetic characterization of 17 congeneric inhibitors of the metalloprotease TLN by SPR and their correlation with structural data reveals insights into the molecular mechanism of binding. We suggest that the involvement of the Asn112 side chain in hydrogen bonds with the bound ligand provokes a prolonged residence time of the respective inhibitor. Asn112 is crucially involved in the motion transforming the protease from its opened to closed state. We therefore propose that a reinforced interaction to Asn112 by the bound ligand restricts the conformational flexibility of this residue. This deteriorates ligand release from the bound state as the Asn112 side chain has to move out of space to allow access to the S_1' pocket. The activation of this retrograde induced-fit step can be further impeded by augmenting the steric hindrance of the Asn112 movement, as observed in the complex of the highly potent TLN inhibitor ZF^{P}LA (**18**). This ligand exhibits a drastic decrease in the dissociation rate constant by a factor of 74,000 compared to its analog **1** [264]. Additionally, a fine-tuning of the binding kinetic properties results from modulations of the P_2' portion of the ligands that are linked to variations in the structure of the surface water network which wraps around the S_2' pocket. Perturbance of these networks of hydrogen-bonded water molecules may lead to an enhanced dissociation rate.

These results highlight the determining role of changes in protein conformation on the binding kinetics of bound ligands. Therefore, a profound understanding of protein flexibility, experienced during the ligand binding event, is of vital importance for a rational design and optimization of ligand binding kinetic parameters.

6.6 Methods

6.6.1 Surface Plasmon Resonance Measurements

SPR measurements were performed on a Biacore T200 system (GE Healthcare) with analysis and sample compartment temperature set to 25 °C. The binding kinetic assay was developed as a capture assay of biotinylated TLN utilizing the Biotin CAPture Kit (GE Healthcare) combined with single cycle kinetics of the inhibitors as previously described [16]. Briefly, sensor Chip CAP was docked in the Biacore and prepared according to the manufacturer's instructions. For each day of SPR measurements, new dilutions of inhibitors from 10 mM stock solutions in DMSO (stored at –20 °C) and biotinylated TLN, were made using freshly prepared buffer. 50 mM Tris pH 8.0, 100 mM NaSCN, 2 mM $CaCl_2$, 2% DMSO was used as sample and running buffer. Runs included three start-up cycles and each analysis cycle contained 4 steps: (1) injection of Biotin CAPture reagent (streptavidin modified with a deoxyriboologonucletide that hybridizes with the complimentary oligonucleotide present on the Sensor Chip CAP) in both reference and active flow cells for 300 s and at 2 µl/min resulting in response levels of around 3000 RU, (2) injection of 150–200 µg mL^{-1} TLN-biotin at 10 µL/min for 90–180 s in active flow cell only resulting in capture levels of 800–1200 RU (a capture stabilization time of 300 s was applied for some runs), (3) injection of inhibitor in increasing concentration over reference and active flow cell using a single cycle kinetics procedure with five 120 s injections at 30 µL min^{-1} and a 240 s dissociation time after the final injection, and (4) injection of standard (6 M guanidine-HCl, 0.25 M NaOH) and additional (30% acetonitrile in 0.25 M NaOH) regeneration solutions to remove the Biotin Cap Reagent, TLN-biotin and any bound inhibitor from both flow cells. Inhibitors were analyzed in duplicate or triplicate measurements using concentrations in the range of 25–2000 nM for **2**, **4**; 25–2500 nM for **16**, **17**; 156–2500 nM for **1**, **3**, **6**, **11**; 156–10000 nM for **8**, **9**, **14**, **15**, 625–10000 nM for **5**, **10**, **12**, **13** and 1250–20000 nM for **7**. Analysis of **2**, **6** and **8** at higher ionic strength was performed in duplicate measurements at concentrations between 156–2500 nM using the above running buffer supplemented with 100 mM and 500 mM NaCl, respectively. Blank cycles defined as analysis cycles with buffer only in step (3), were performed first, last and between every inhibitor concentration series. Data were double referenced by first subtraction of reference flow cell and then subtraction of blank cycles. Fitting of data was performed using Biacore T200 evaluation software 3.0, applying a 1:1 binding model compensating for linear drift.

6.6.2 Crystal Preparation and Soaking

Lyophilized TLN powder (Calbiochem) was dissolved in 50 µL DMSO to yield a solution with a concentration of 8 mM, followed by the addition of 50 µL of a solution containing 3.8 M CsCl and 100 mM Tris-HCl (pH 7.5). After centrifugation (3 min, 16,000 g), 1 µL of the clear

solution was pipetted into the wells of a 24 well sitting drop crystallization plate (reservoir wells filled with 1 mL demineralized water). The crystal plate was sealed and crystals finished growing after 5 days at 18 °C. Ligands were soaked into the crystals at a concentration of 1 mM for 24 hrs in a buffer composed of 100 mM Tris-HCl (pH 7.5), 2 mM $CaCl_2$, 5% DMSO and subsequently flash-frozen in liquid nitrogen in a buffer composed of 10 mM Tris-HCl (pH 7.5), 10 mM $Ca(CH_3COO)_2$, 5% DMSO, 20% glycerol and 1 mM of the respective ligand.

6.6.3 Data Collection, Processing, Structure Determination and Refinement

Data collection of the crystal structures TLN-**5** and TLN-**6** (Supporting Information Table S6.3) was performed at BESSY II (Helmholtz-Zentrum Berlin) at the MX-beamline 14.1. Datasets were collected on a Dectris Pilatus 6M pixel detector at a wavelength of 0.91841 Å. *XDS* [158] was used for indexing, integration and scaling of the datasets. The structures were determined by molecular replacement using *Phaser* (version 2.5.0) [159] applying the structure with the PDB code 8TLN [160] as a search model. Subsequently, model refinement (xyz coordinates, individual *B* factors, occupancies) with *Phenix.refine* (version 1.10.1-2155) [162] and model building into σA-weighted maps ($2F_o$–F_c and F_o–F_c) with *Coot* (version 0.7) [186] were performed in alternating cycles until *R* values reached convergence. A randomly chosen subset of 5% of the reflections was excluded from the refinement and used for the calculation of R_{free}. As a first refinement step, Cartesian simulated annealing was performed (default settings). *B* factors for all model atoms (except for hydrogen atoms) were refined anisotropically. Hydrogen atoms (riding model) were added to the amino acids with *Phenix.refine*. Alternative conformations of amino acid side chains and ligand moieties were assigned to the electron density if an occupancy of at least 20% was obtained after refinement. Ligand molecules were modeled with *MOE* [269] and restraints were created with *Phenix.elbow* [163] (**5**) or with the *Grade* web server [270] (**6**).

6.6.4 Accession Codes

Atomic coordinates and experimental details for the crystal structures of **5** and **6** (PDB codes 5LIF and 5LWD) will be released upon publication.

6.7 Acknowledgement

The authors want to thank the MX-team at BESSY II (Helmholtz-Zentrum Berlin, Germany) for their advice during data collection. The authors acknowledge the receipt of a travel grant from the Helmholtz-Zentrum Berlin. This work was funded by the European Research Council

(ERC) of the European Union (grant 268145-DrugProfilBind). TLN inhibitors **1**–**5** and **7**–**13** were kindly provided by N. Nasief and. D. G. Hangauer (Univ. Buffalo, NY, USA).

6.8 Supporting Information

6.8.1 Experimental data from SPR measurements

Table S6.1. SPR derived kinetic data of thermolysin inhibitors **1**–**17**.

Compound	k_a [$M^{-1}s^{-1}$]	SD k_a [$M^{-1}s^{-1}$]	log k_a	k_d [s^{-1}]	SD k_d [s^{-1}]	log k_d	K_d [μM]
1	2.38×10^4	3.54×10^3	4.38	5.06×10^{-3}	1.63×10^{-4}	-2.30	0.21
2	6.88×10^4	6.46×10^2	4.84	2.77×10^{-3}	8.49×10^{-5}	-2.56	0.04
3	3.53×10^4	2.40×10^3	4.55	2.87×10^{-3}	3.04×10^{-4}	-2.54	0.08
4	4.94×10^4	6.93×10^3	4.69	7.86×10^{-3}	4.31×10^{-4}	-2.10	0.16
5	3.47×10^4	6.01×10^3	4.54	4.63×10^{-2}	1.02×10^{-3}	-1.33	1.34
6	7.88×10^4	7.07×10^3	4.91	4.46×10^{-2}	3.25×10^{-3}	-1.35	0.57
7	1.88×10^5	1.34×10^4	5.27	6.15×10^{-1}	4.95×10^{-3}	-0.21	3.28
8	1.44×10^5	9.19×10^3	5.16	1.71×10^{-1}	7.07×10^{-4}	-0.77	1.19
9	8.96×10^4	1.41×10^3	4.95	2.34×10^{-1}	5.66×10^{-3}	-0.63	2.61
10	1.01×10^5	1.41×10^3	5.00	2.08×10^{-1}	7.78×10^{-3}	-0.68	2.05
11	1.10×10^5	8.06×10^3	5.04	7.94×10^{-2}	3.97×10^{-3}	-1.10	0.72
12	1.01×10^5	1.27×10^4	5.00	1.82×10^{-1}	5.66×10^{-3}	-0.74	1.80
13	8.65×10^4	4.95×10^2	4.94	1.40×10^{-1}	1.27×10^{-2}	-0.85	1.62
14	2.69×10^4	7.08×10^3	4.43	3.87×10^{-2}	5.09×10^{-3}	-1.41	1.44
15	4.00×10^4	1.16×10^4	4.60	7.00×10^{-2}	5.98×10^{-3}	-1.15	1.75
16	8.25×10^4	3.32×10^2	4.92	5.97×10^{-2}	3.07×10^{-3}	-1.22	0.72
17	9.55×10^4	3.32×10^4	4.98	6.64×10^{-2}	1.49×10^{-2}	-1.18	0.70

6.8.2 Experimental data from SPR measurements at varying ionic strength conditions

Table S6.2. SPR derived kinetic data of selected inhibitors at varying ionic strength conditions.

Compound	Ionic strength[a] [mM]	k_a [$M^{-1}s^{-1}$]	SD k_a [$M^{-1}s^{-1}$]	log k_a	k_d [s^{-1}]	SD k_d [s^{-1}]	log k_d	K_d [μM]
2	134	6.88×10^4	3.46×10^3	4.84	2.77×10^{-3}	8.49×10^{-5}	-2.56	0.04
2	234	5.73×10^4	6.06×10^3	4.76	3.29×10^{-3}	2.45×10^{-4}	-2.48	0.06
2	634	4.24×10^4	1.91×10^3	4.63	2.85×10^{-3}	7.25×10^{-4}	-2.55	0.07
6	134	7.88×10^4	7.07×10^3	4.90	4.46×10^{-2}	3.25×10^{-3}	-1.35	0.57
6	234	4.78×10^4	1.19×10^4	4.89	3.38×10^{-2}	3.13×10^{-3}	-1.47	0.43
6	634	5.15×10^4	1.30×10^4	4.71	4.80×10^{-2}	1.85×10^{-3}	-1.32	0.93
8	134	1.44×10^5	9.19×10^3	5.16	1.71×10^{-1}	7.07×10^{-4}	-0.77	1.19
8	234	2.75×10^5	3.00×10^4	5.44	1.93×10^{-1}	3.75×10^{-3}	-0.71	0.70
8	634	2.49×10^5	2.40×10^3	5.40	1.89×10^{-1}	7.64×10^{-3}	-0.72	0.76

[a] Ionic strength was calculated from all ionized buffer components at pH 8.0 as $I = \sum_i c_i \cdot z_i^2$

6.8.3 Sensograms from SPR measurements

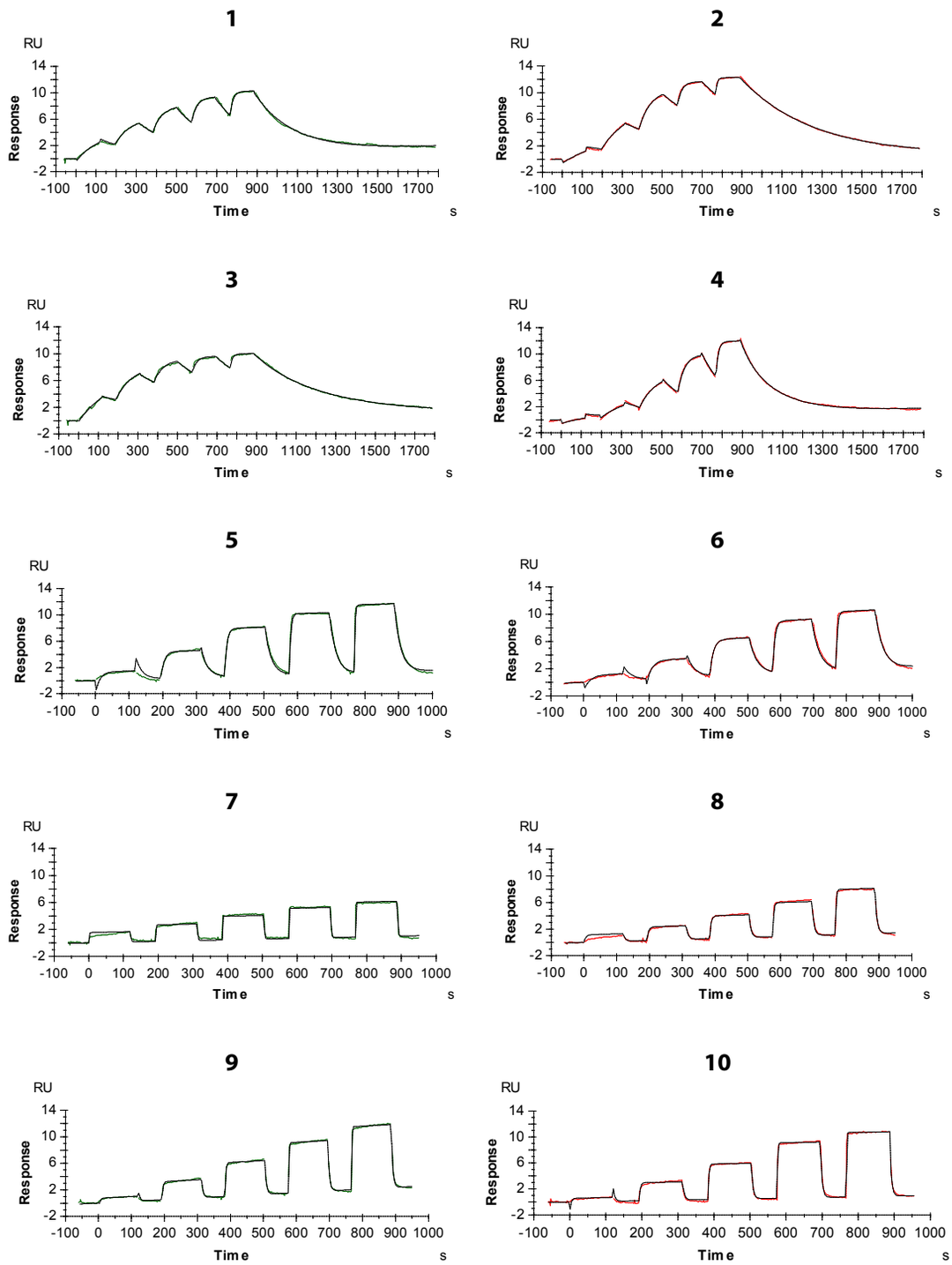


Figure S6.1. SPR sensorgrams showing one representative single cycle kinetics run (green or red) of ligand **1-17**, respectively**,** fitted to a 1:1 kinetic binding model (black). The five concentrations used in the shown sensorgrams were 0.025, 0.074, 0.22, 0.67, 2.0 µM for ligand **2, 4, 16, 17**; 0.16, 0.31, 0.63, 1.25, 2.5 µM for ligand **1, 3, 6, 8, 11**; 0.63, 1.25, 2.5, 5.0, 10.0 µM for ligand **5, 9, 10, 12, 13, 14, 15** and 1.25, 2.5, 5.0, 10.0, 20.0 µM for ligand **7.**

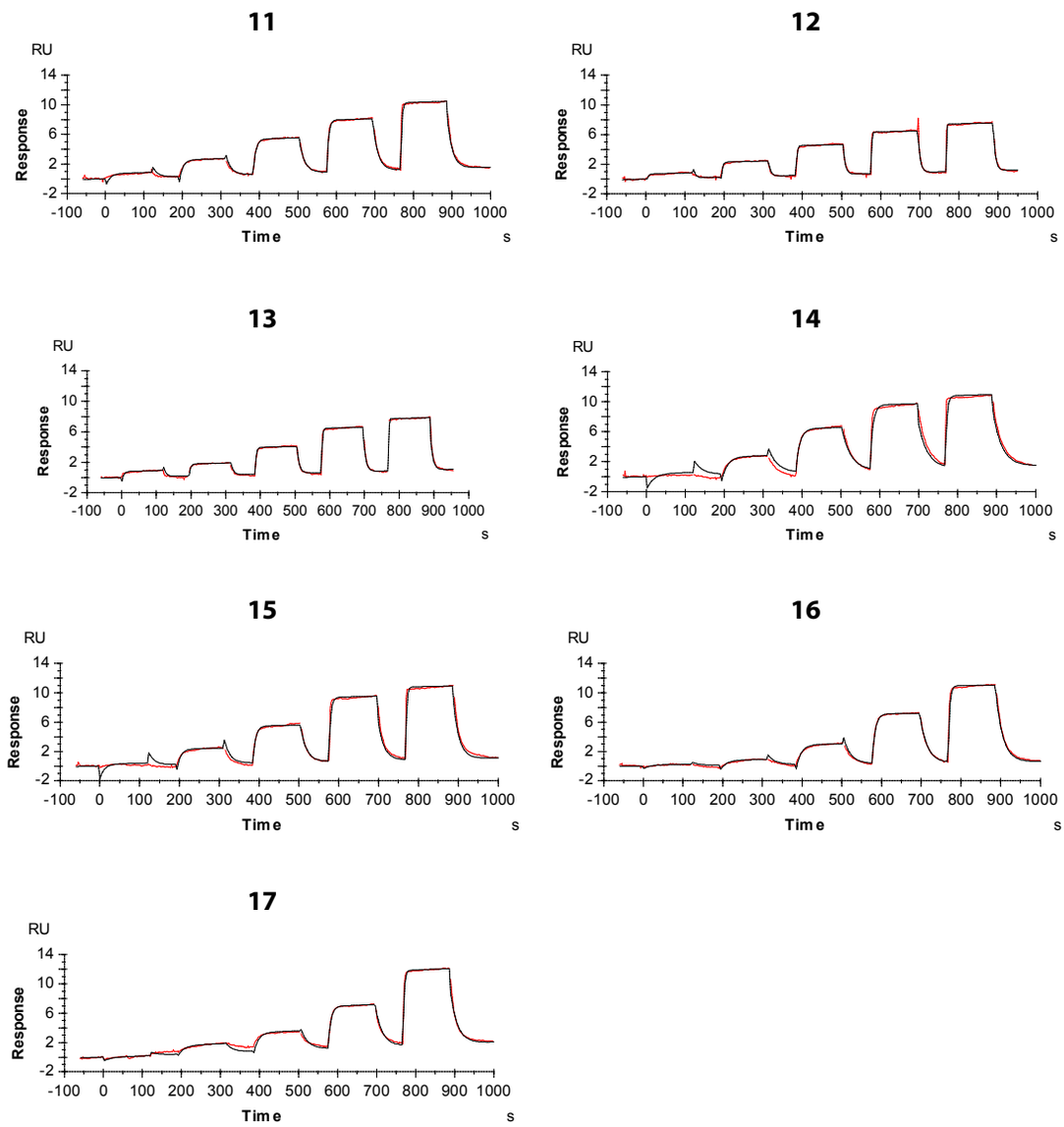


Figure S6.1. (continued)

6.8.4 Comparison of kinetic data from SPR and photometric inhibition assay

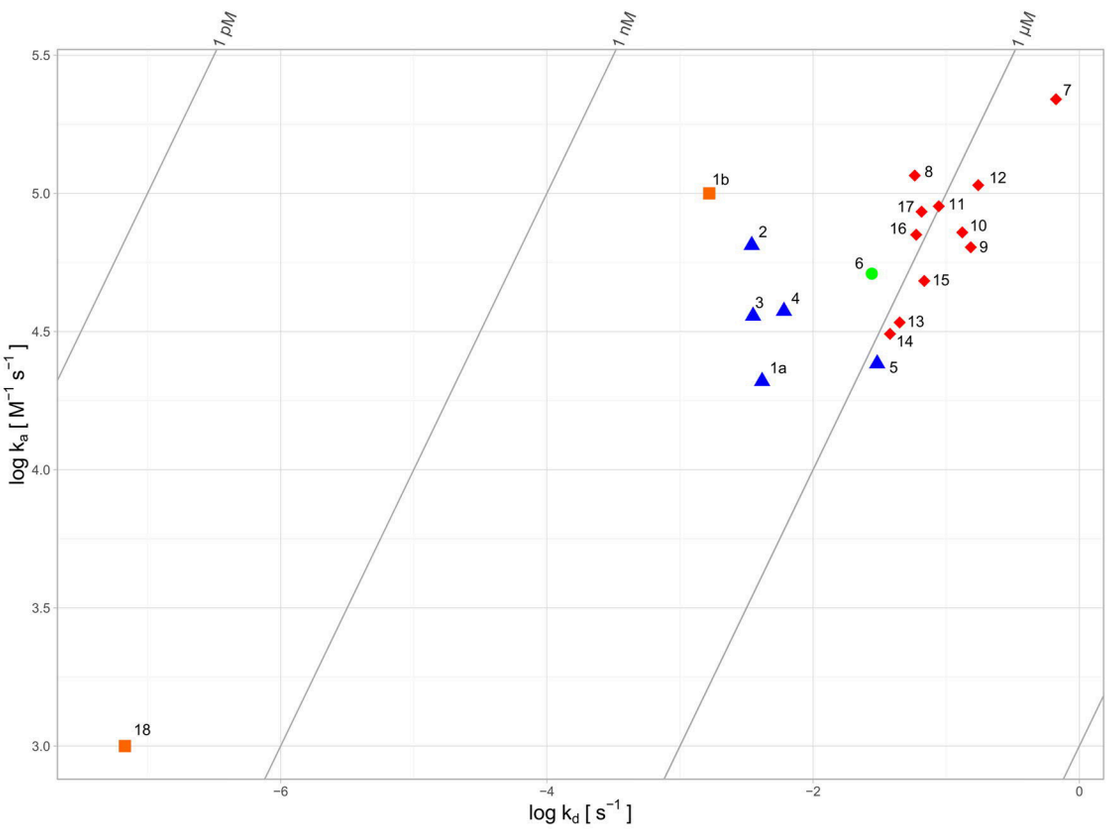


Figure S6.2. Kinetic Map containing binding kinetic data of **1b** and **18** taken from Bartlett and Marlowe (orange blocks) [264]. From the kinetic rate constants of **1a** and **1b** the variation between the inherently different experimental methods (SPR vs. photometric inhibition assay) can be estimated to fall within one order of magnitude.

6.8.5 **Crystallographic tables**

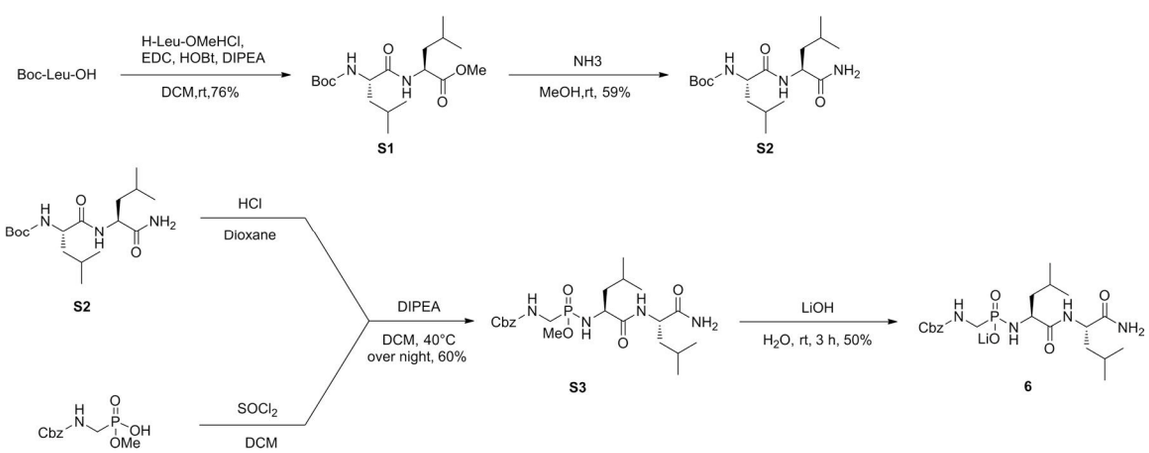
Table S6.3. Data collection and refinement statistics for crystal structures 5LIF and 5LWD.

	Complex (PDB code)	
	TLN-**5** (5LIF)	TLN-**6** (5LWD)
Data collection and processing		
Space group	$P6_122$	$P6_122$
Unit cell parameters: *a*, *b*, *c* (Å)	91.8, 91.8, 130.0	92.8, 92.8, 130.6
Matthews coefficient (Å^3/Da)[a]	2.3	2.4
Solvent content (%)[a]	46.3	47.7
Resolution range (Å)	50.00-1.31 (1.31–1.39)	50.00-1.23 (1.30–1.23)
Wilson *B* factor (Å^2)	10.1	9.1
Unique reflections	77996 (12256)	96448 (15229)
R_{sym} (%)	5.7 (49.0)	6.4 (47.3)
Completeness (%)	99.8 (99.0)	99.7 (98.8)
Redundancy	10.6 (10.3)	24.9 (24.6)
$<I/\sigma(I)>$	27.3 (5.2)	36.3 (7.3)
Refinement		
Resolution range (Å)	19.66–1.31	31.75–1.23
Reflections used in refinement (work/free)	74096/3900	91621/4822
R_{cryst} (%)	11.3	10.4
R_{free} (%)	13.6	12.3
Protein residues	316	316
Calcium/zinc ions	4/1	4/1
Inhibitor atoms	35	32
Water molecules	407	401
RMSD from ideality:		
Bond lengths (Å)	0.008	0.010
Bond angles (°)	1.0	1.2
Ramachandran plot (%):[b]		
Residues in most favored regions	87.8	88.5
Residues in additionally allowed regions	11.1	10.4
Residues in generously allowed regions	0.7	0.7
Residues in disallowed regions[c]	0.4	0.4
Mean *B* factor (Å^2):[d]		
Protein	10.5	10.7
Inhibitor	12.5	10.7
Water molecules	26.0	26.6

The highest resolution shell is described by values in parentheses. [a]Matthews coefficient and solvent content were calculated with the program *Matthews_coef* from the *CCP*4 suite (version 6.3.0) [180]. [b]Ramachandran plots were calculated with *PROCHECK* [138]. [c]The Ramachandran outlier (Thr26) occurs in every structure of TLN and is described in literature [181]. [d]Mean *B* factors were calculated with *MOLEMAN* [188].

6.8.6 Ligand synthesis and purification

Experimental details. ^{1}H, ^{13}C and ^{31}P NMR spectra were recorded on a JEOL ECX-400 or JEOL ECA-500 instrument. All chemical shift values are reported in ppm relative to the non-deuterated solvent signal. An external standard was used for ^{31}P NMR spectra (referenced to: 85% H_3PO_4) and ^{13}C NMR spectra in D_2O (referenced to: trimethylsilyl propanoic acid). For the description of multiplicity the following abbreviations were used: s = singlet, m = multiplet. ESI-MS spectra were recorded on a Q-Trap 2000 system by Applied Biosystems. For high resolution ESI-MS a LTQ-FT Ultra mass spectrometer (Thermo Fischer Scientific) was used. For HPLC chromatography a Shimadzu LC-20 system equipped with a diode array detector was used. Analytic separations were carried out with a MN Nucleodur 100-5 C18 ec 4.6×250 mm column using a water-acetonitrile gradient. For semi-preparative separations a Water XSelect CSH C18 10×250 mm column employing a water-acetonitrile gradient was used.



Scheme S6.1. Synthesis of compound **6** from Boc-Leu-OH.

Boc-Leu-Leu-OMe (**S1**). To a suspension of Boc-Leu-OH (347 mg, 1.50 mmol, 1.0 eq), EDC (374 mg, 1.95 mmol, 1.3 eq), HOBt (264 mg, 1.95 mmol, 1.3 eq), and H-Leu-OMe HCl(272 mg, 1.50 mmol, 1.0 eq) in DCM, DIPEA (485 mg, 3.75 mmol, 2.5 eq) was added and the reaction mixture was stirred at rt overnight. The solvent was removed under reduced pressure and the resulting residue was taken up in EtOAc and extracted with 1 M HCl (3 × 10 mL) and sat. $NaHCO_3$ (3 × 10 mL). The organic phase was washed with brine and dried over anhydrous $MgSO_4$. The crude product was purified by silica gel chromatography (DCM/MeOH 50:1) to give Boc-Leu-Leu-OMe as a colorless solid (410 mg, 1.14 mmol, 76%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.78 – 1.00 (m, 12H), 1.43 (s, *J*=9.8, 9H), 1.58 – 1.75 (m, 6H), 3.72 (s, 3H), 4.03 – 4.15 (m, 1H), 4.61 (td, *J*=8.7, 4.7, 1H), 4.86 (d, *J*=7.7, 1H), 6.43 (d, *J*=8.2, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 21.9, 22.2, 22.9, 24.7, 28.3, 40.9, 41.6, 50.7, 52.4, 53.0, 80.2, 155.8, 172.4, 173.3. **MS (ESI+)** *m/z* calculated for $C_{18}H_{35}N_2O_5$ $[M+H]^+$: 349.49; found: 359.24.

Boc-Leu-Leu-NH_2 (**S2**). To 7 mL of a 7 M solution of NH_3 in MeOH, Boc-Leu-Leu-OMe (427 mg, 1.19 mmol) was added. The mixture was stirred at rt for 48 h, the solvent was removed under reduced pressure and the crude product was purified by silica gel chromatography (cyclohexane/EtOAc 1:1). The product was isolated as a colorless solid (239 mg, 0.70 mmol, 59%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.85 – 0.99 (m, 12H), 1.43 (s, 9H), 1.45 – 1.83 (m, 6H), 4.06 (bs, 1H), 4.41 – 4.53 (m, 1H), 4.93 (s, 1H), 5.54 (s, 1H), 6.47 – 6.64 (m, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 21.8, 22.0, 23.0, 23.2, 24.8, 28.4, 40.7, 40.9, 51.3, 53.6, 80.5, 156.1, 175.0, 191.2. **MS (ESI+)** *m/z* calculated for $C_{17}H_{34}N_3O_4$ $[M+H]^+$: 344.48; found: 344.21.

*Phosphonamidate **S3***. $SOCl_2$ (238 mg, 2.00 mmol, 4.0 eq) was dissolved in 5 mL dry DCM under argon and cooled to 0 °C. A solution of benzyl ((hydroxy(methoxy)phosphoryl) methyl)carbamate (233 mg, 0.90 mmol, 1.8 eq) was added to the solution over 30 minutes. The cooling bath was removed and the reaction was stirred at rt for 3–5 h. All volatile components were removed under reduced pressure. In a second flask, Boc-Leu-Leu-NH_2 (172 mg, 0.50 mmol, 1.0 eq) was treated with HCl (4 M in dioxane, 0.5 mL, 2 mmol, 4.0 eq) for 1 h. All volatile components were removed under reduced pressure. The activated phosphonic acid was dissolved in dry DCM. DIPEA and the deprotected dipeptide were added to the solution consecutively. The mixture was heated to 40 °C and stirred overnight. The solution was diluted with EtOAc and extracted with 5% citric acid (3×10 mL), 1 M HCl (1×10 mL) and 1 M NaOH (3×10 mL). The organic phase was washed with brine and dried over $MgSO_4$. The product was obtained as a mixture of diastereomers in form of a brown solid (145 mg, 0.30 mmol, 60%). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 0.71 – 1.06 (m, 12H), 1.16 – 2.27 (m, 6H), 3.40 – 3.82 (m, 5H), 3.81 – 4.00 (m, 1H), 4.50 (s, 1H), 5.00 – 5.23 (m, 2H), 7.25 – 7.49 (m, 5H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 21.7, 21.85, 21.90, 22.1, 23.0, 24.5, 24.6, 24.85, 24.94, 37.4, 37.8, 38.7, 39.3, 41.0, 41.1, 43.7, 43.9, 51.3, 51.4, 52.0, 53.4, 53.5, 67.15, 67.24, 128.1, 128.3, 128.6, 136.28, 136.36,

156.8, 157.0, 174.2, 174.3, 175.7, 175.9. **^{31}P NMR** (162 MHz, $CDCl_3$) δ = 28.6, 29.7. **MS (ESI+)** calculated for $C_{22}H_{37}N_4NaO_6P$: 507.52 $[M+Na]^+$; found: 507.42.

*Inhibitor **6**:* The protected phosphonamidate was treated with 3 mL of a 0.4 M solution of LiOH in water. MeCN was added until the solution cleared. The reaction was stirred for 4 h at rt. Under ice cooling the pH was adjusted to 8 using 5% AcOH and the solvent was removed under reduced pressure. The residue was dissolved in 1 mL of water and purified by semi-preparative HPLC. Lyophilization of fractions containing the pure product afforded a colorless solid (36 mg, 0.08 mmol, 50%). **^{1}H NMR** (500 MHz, D_2O) δ = 0.89 – 1.01 (m, 12H), 1.46 – 1.59 (m, 2H), 1.63 – 1.84 (m, 4H), 3.19 – 3.58 (m, 2H), 3.65 – 3.85 (m, 1H), 4.29 – 4.44 (m, 1H), 5.09 – 5.28 (m, 2H), 7.38 – 7.66 (m, 5H). **^{13}C NMR** (101 MHz, D_2O) δ = 20.1, 21.0, 22.3, 22.4, 24.2, 24.4, 39.1 (d, J_{CP} = 134.1 Hz), 39.3, 42.9, 51.8, 54.5, 67.2, 127.8, 128.4, 128.8, 136.4, 158.2, 178.1, 178.9. **^{31}P NMR** (202 MHz, D_2O) δ = 19.1. **HRMS (ESI+)** calculated for $C_{21}H_{35}N_4O_6PNa$: 493.2186 $[M+Na]^+$; found: 493.2184.

6.8.7 Sequence alignment of representative M4 and M13 proteases

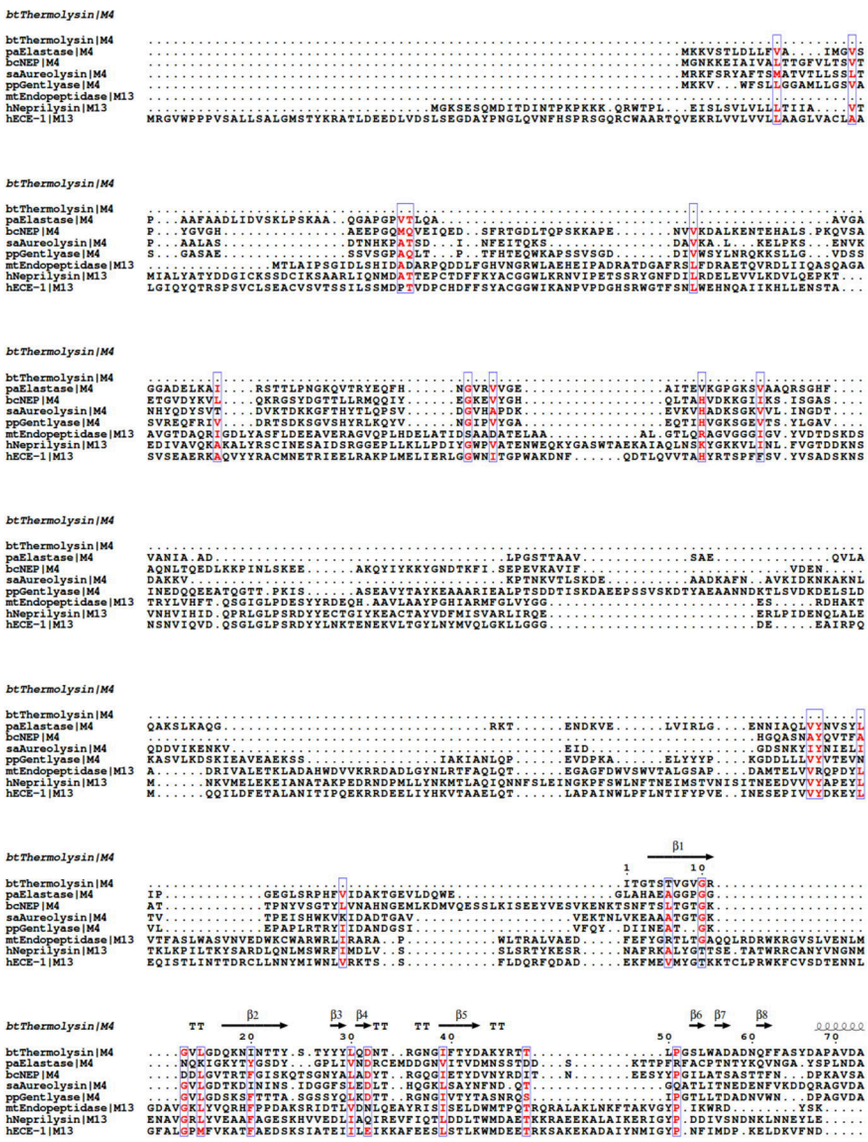


Figure S6.3. Secondary structure alignment of thermolysin and proteases from the M4 and M13 family. Only the main chain of thermolysin (Uniprot P00800) was used for the alignment. The sequences of the signal- and propeptide were omitted to conserve canonical numbering throughout the main text. The protein structure comparison service PDBeFold at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm) was used for secondary structure alignment [298]. The results were displayed using the ESPript server (http://espript.ibcp.fr) [299].

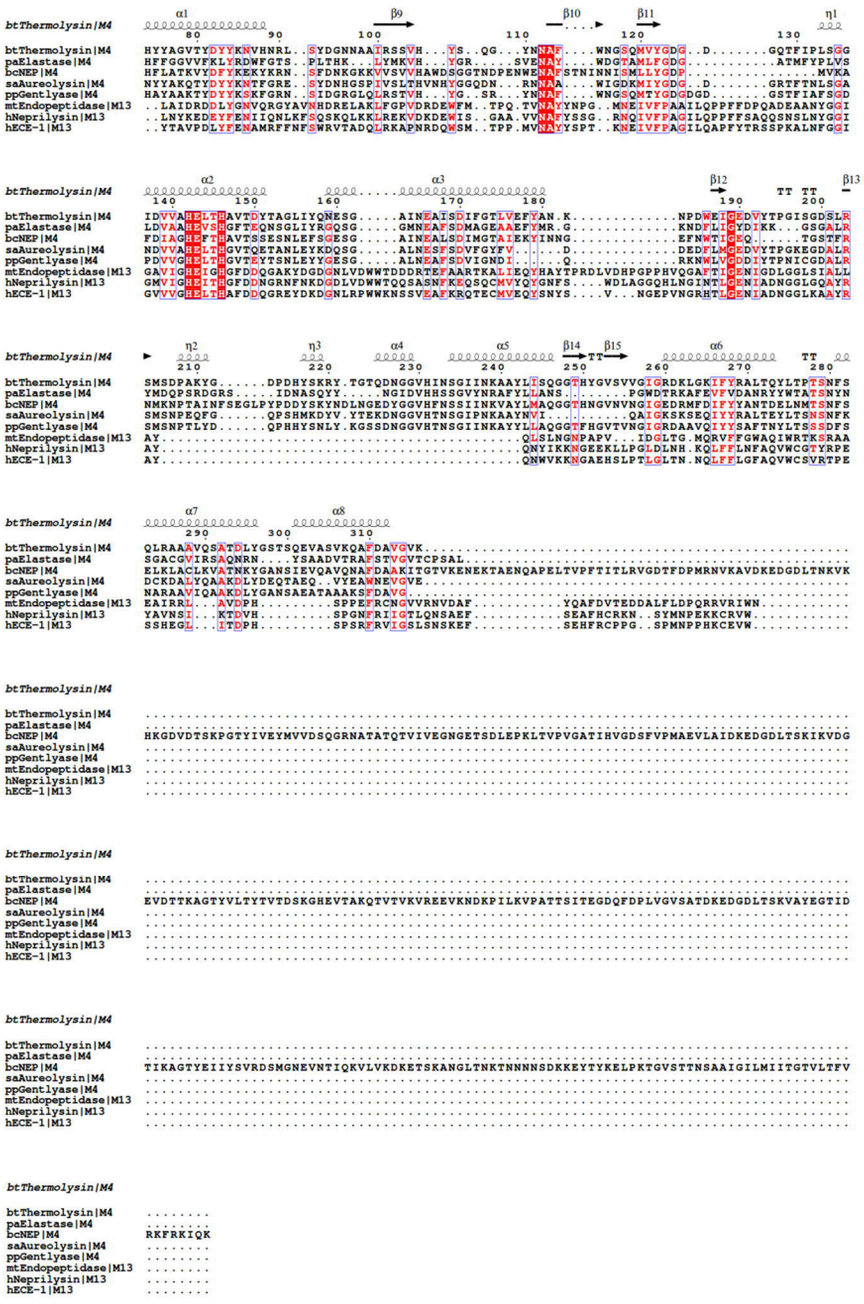


Figure S6.3. Continued.

Chapter 7

Watch Out for the Red Herring: Surprising Reactivity of a Fragment Results in Biological Activity

Jonathan Cramer[1,#], Dr. Johannes Schiebel[1,#], Tobias Wulsdorf[1], Kristof Grohe[2,3], Eszter Eva Najbauer[2], Frederik R. Ehrmann[1], Nedyalka Radeva[1], Nina Zitzer[1], Dr. Uwe Linne[4], Prof. Dr. Rasmus Linser[2,3], Prof. Dr. Andreas Heine[1], Prof. Dr. Gerhard Klebe[1]

[#]These authors contributed equally

[1]Institut für Pharmazeutische Chemie, Philipps-Universität Marburg
Marbacher Weg 6, 35032 Marburg (Germany)

[2]Abteilung für NMR-basierte Strukturbiologie, Max Planck Institut für Biophys. Chemie
Am Faßberg 11, 37077 Göttingen (Germany)

[3]Fakultät für Chemie und Pharmazie, Ludwig-Maximilians-Universität München
Butenandtstr. 5-13, Haus F, 81377 München (Germany)

[4]Fachbereich Chemie, Philipps-Universität Marburg
Hans-Meerwein-Straße 4, 35043 Marburg (Germany)

Angewandte Chemie International Edition **2017**, 7: 1908–1913

Angewandte Chemie **2017**, 6: 1934–1940

7.1 Abstract

With the rising popularity of fragment-based approaches in drug development, more and more attention has to be devoted to the detection of false-positive screening results. In particular, the small size and low affinity of fragments drives screening techniques to their limit. The pursuit of a false-positive hit can cause significant loss of time and resources. Here, we present an instructive and intriguing example about the origin of misleading assay results for a fragment that emerged as most potent binder for the aspartic protease endothiapepsin (EP) across multiple screening assays. This molecule shows its biological effect mainly after conversion to another entity through a reaction cascade that involves major rearrangements of its heterocyclic scaffold. The formed ligand binds EP through an induced-fit mechanism involving remarkable electrostatic interactions. Structural information in the initial screening proved to be crucial for the identification of this false-positive hit.

Keywords

drug discovery • fragment-based lead discovery • PAINS • reactivity • medicinal chemistry

7.2 Main Text

Fragment-based approaches are very efficient in the generation of new lead molecules required for the design of drug candidates. However, the screening methods that are commonly applied to identify promising hits for subsequent X-ray crystallographic experiments can give misleading results. The identification of false-positives, often caused by so called Pan Assay Interference Scaffolds (PAINS) [271], is a major challenge in any library screening program. It has been shown that the molecular mechanisms for the generation of false-positive signals include unspecific binding, colloidal aggregation, inherent reactivity and interference with the assay detection method [81,272–274]. Additionally, the presence of impurities from synthesis or degradation of the compound can give misleading results [68,76,275]. The pursuit of a false-positive binder can lead to a significant loss of time and resources. Therefore, the identification of such "red herrings" has become a major concern in library design.

Recently we performed an extensive screening of a fragment library containing 361 compounds against the aspartic protease endothiapepsin (EP) [64,65]. In this context we compared the outcome of various state-of-the-art screening methods. Additionally, the entire library was subjected to a crystallographic screen with a subsequent characterization of the hits by isothermal titration calorimetry (ITC). Compound **1** (fragment number 177) emerged as one of the most potent binders across multiple screening methods. It was found among the top ten ranking fragments in four out of six and as hit in five out of six screening techniques, showing virtually complete inhibition of the enzyme in two different biochemical assays, the highest thermal shift of 3.4 ± 0.3 °C and a K_d of 115 ± 8 μM by ITC [64,65]. Importantly, during the selection of molecules for the STD-NMR experiments, the fragment passed a filter process based on purity, aggregation behavior and solubility criteria typically applied in pharmaceutical companies [65]. Moreover, **1** is not predicted as PAINS using the approach of Baell and Holloway [81]. Therefore **1** seemed to be an attractive starting point for a fragment-based lead optimization campaign.

With our established crystallographic protocol, we were able to collect a dataset of **1** in complex with EP to a resolution of 1.25 Å. To our surprise the ligand, which was bound near the active site of the enzyme, showed little resemblance to **1** in four independent crystallization experiments. The electron density clearly indicated the presence of a much larger molecule accommodated in the binding cleft (Figure 7.1A). This molecule featured a tricyclic core that was substituted with a bicyclic and a monocyclic moiety. The characteristic methyl substitution pattern and the unambiguous presence of a chlorine atom, as indicated by its anomalous scattering, suggested a chemical relation to **1**. Fortunately, the high resolution of the crystallographic model, referred to as EP-**2** in the following, allowed us to determine the atomic connectivity of the unknown binder **2**.

Supplier specifications as well as our own HPLC and NMR experiments excluded the possibility of an impurity in the sample that could account for the crystallographic results. Therefore, we proposed that **2** has to be generated from **1** under the applied assay conditions. To verify this hypothesis, we monitored a solution of **1** in the assay buffer by HPLC, which clearly showed the time-dependent formation of multiple species in the reaction mixture. Since some material precipitated from aqueous solution, we repeated the experiment in methanol at 50 °C, which gave equivalent results (Figure S7.1A). Mass spectrometric analysis revealed molecules with m/z values of 192.3, 196.1, 337.3, 355.3 and 533.1 as the main components of the mixture. Subsequently, these compounds were isolated by preparative HPLC and subjected to HR-MS, MS/MS, NMR and protein X-ray crystallography, leading to the assignment of putative structures (Scheme 7.1).

m/z 192.3: MS analysis indicated the absence of a chlorine substituent and thus allowed the identification of **3**, the alcoholysis product of **1**. This finding suggests that the chloropyridazine heterocycle in **1** is prone to react in a nucleophilic substitution reaction. Remarkably, the methoxy derivative **3** was found to bind to EP. Four copies of the molecule, two of which had sufficient occupancy to assign them to the electron density in the crystal structure, bind upon Phe291, which is known to be a binding hotspot in EP [67], in a 5-fold stacking arrangement (Figure S7.2A, structure referred to as EP-**3**).

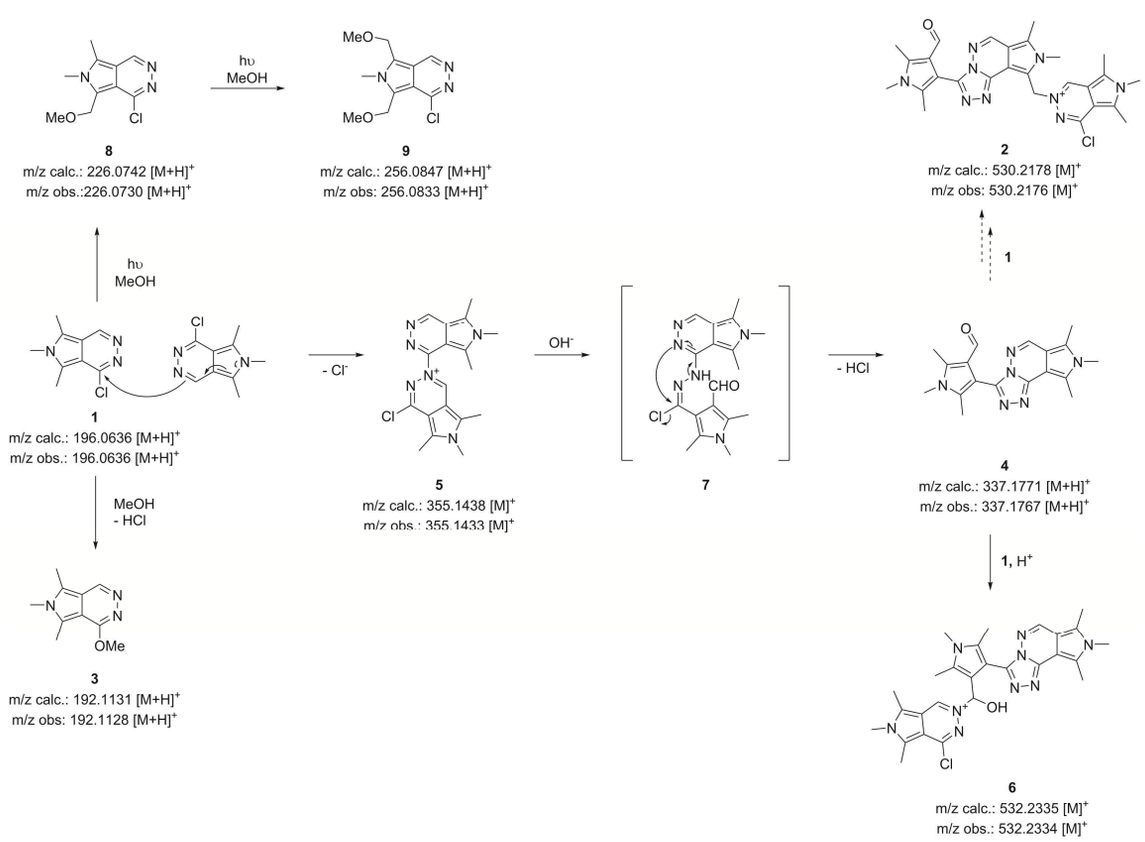
m/z 196.1: HR-MS and NMR results confirmed that this molecule corresponds to the parent compound **1**. Furthermore, we could reproduce the above-mentioned HPLC reaction pattern by a repeated exposition of the isolate to methanol at 50 °C. This observation verifies the reactive nature of **1** in solution. Soaking experiments with this isolate resulted in a structure containing **2** with some density for additionally bound **1**, now visible in comparison to EP-**2**, presumably due to the higher concentration or resolution (Figure 7.1B, structure referred to as EP-**1**-**2**). It could therefore be excluded that trace impurities from the synthesis of **1** accumulate in the binding pocket of EP or that such impurities are required for the formation of **2**.

m/z 337.3: From the mass spectrum it was obvious that no chlorine atoms were present in this molecule. In agreement with our HR-MS and NMR experiments, we assigned structure **4** to this compound. Importantly, this structure closely resembles the central scaffold of **2**, the putative molecule initially found in the binding pocket of EP.

m/z 355.5: The isotopic distribution of this molecule suggested the presence of a single chlorine atom, leading to the assignment of structure **5**. NMR, HR-MS and MS/MS analyses supported this proposal (Scheme S7.1). Crystallographic experiments did not reveal electron density in the EP active site that might correspond to **5**. Instead, density for parts of **4** could be detected (Figure S7.2B, structure referred to as EP-**4**), which might be explained by the finding that the better water-soluble **5** is readily converted into **4**.

m/z 533.1: The associated NMR spectrum indicated the presence of a trimeric form of **1**. However, the absence of a CH_2-group excluded a composition that could explain the electron density observed in the crystal structure. Instead, in agreement with HR-MS and MS/MS experiments (Scheme S7.1), structure **6** could be assigned to this molecule. Consistently, **6**, which we could provide for interaction studies after HPLC isolation, did not bind to EP in crystallographic experiments.

From the chemical composition of the isolated compounds, we deduced a mechanism for the reaction of **1** in methanolic solution, which is depicted in Scheme 7.1. The reaction pathway is initiated by a nucleophilic attack of an imino nitrogen of **1** on the 3-chloropyridazine ring of a second molecule. This dimerization results in the formation of the charged intermediate **5**. In a competitive reaction, the halogen atom is exchanged against a methoxy group upon nucleophilic attack of a solvent molecule, which forms **3**. After its generation, **5** is transformed into **4** by an intramolecular substitution reaction. This reaction is proposed to proceed via the hypothetical intermediate **7** that is formed by a hydrolytic cleavage of the iminium moiety in **5**. A similar reaction cascade has been described for unsubstituted 6H-pyrrolo[3,4-*d*]



Scheme 7.1. Proposed reaction mechanism of **1** in methanolic solution. Experimentally observed HR-MS data are compared to calculated m/z values of the postulated reaction intermediates.

pyridazines [276]. The suggested reaction pathway could be supported by the treatment of **5** with an aqueous ammonia solution. In this experiment, HR-MS analysis showed a quantitative transformation of **5** to **4**. Under low-pH conditions, however, a transformation of **5** to **4** is severely decelerated. Compound **4** finally reacts with an additional molecule of **1** to the trimeric form **6**, a charged hemi-aminal species.

To explain the inherent reactivity of **1**, we calculated *Nucleus Independent Chemical Shifts* (NICS), natural atomic charges, Wiberg bond indices and electrophilicity indices for this molecule and several derivatives thereof using QM methods at a high level of theory (Table S7.2). The aromatic nature of the compound's pyridazine ring has been investigated by a comparison with related isoindole, [2',3']-pyrrolo annelated indole-like and monocyclic analogs (Table S7.2). The NICS(1) calculations (Table S7.11) revealed a loss in aromaticity on the six-membered ring of **1** and other isoindoles compared to indole analogs. At the same time, an increased bond ordering for the C=N bond in **1** compared to pyridazine and its [2',3']-pyrrolo annelated indole-like analogs is indicative for an imine-like C=N bond, facilitating the

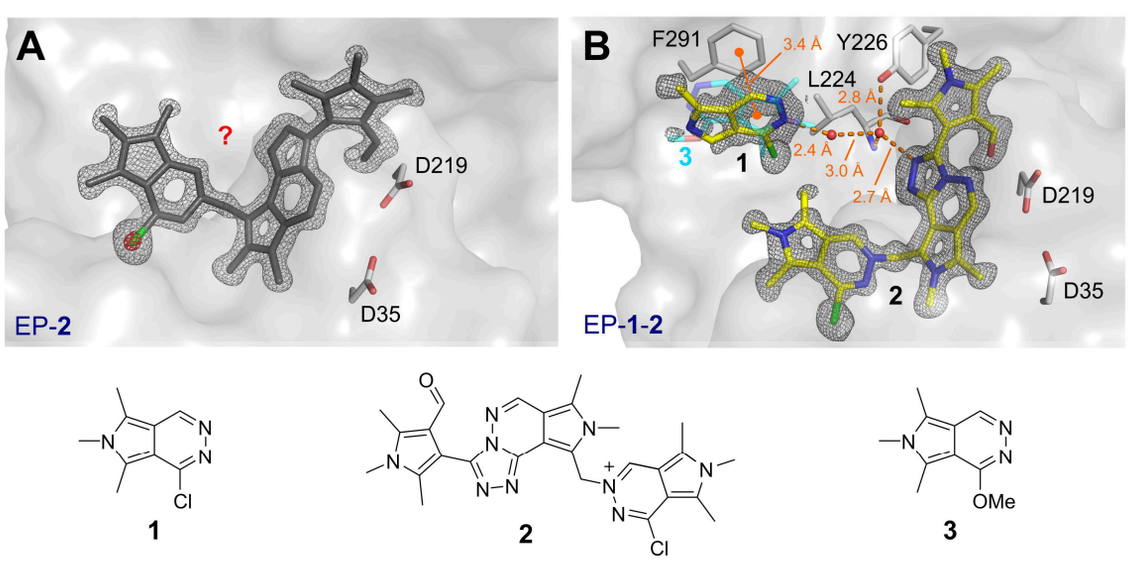


Figure 7.1. Crystal structures of EP in complex with different reaction products of **1** (see also Figure S7.3). The mF_o-DF_c electron density for bound ligands are depicted as gray meshes at the 3σ level (2.8σ for **1** in panel B) prior to the inclusion of each respective ligand into the model. (A) Structure of an unknown contaminant in complex with EP from the crystallographic screening campaign (EP-**2**). The observed electron density clearly cannot be explained by the chemical formula of **1** but instead reveals the presence of a larger molecule that is characterized by the connectivity defined via gray sticks and contains a chlorine atom as indicated by the anomalous signal shown in red (5σ level). (B) Structure generated by soaking of an EP crystal with the HPLC isolate m/z = 196.1 (EP-**1**-**2**). The specific interaction of **1** with its environment is highlighted in orange while the methoxy analog of **1** (molecule **3**) is shown with its primary binding mode from a superimposition of the EP-**3** onto the EP-**1**-**2** structure in transparent cyan (see also Figure S7.3A).

initial reaction. The high global electrophilicity index of **1** (Table S7.16) and the positive atomic charge of the carbon atom attached to the chlorine (Table S7.3) further indicate that this position is prone to nucleophilic attacks. A putative positive charge, as proposed for compound **2**, should lead to an additional gain in reactivity, as indicated by increased electrophilicity and decreased HOMO-LUMO gap (**P-4** in Table S7.15, Table S7.16). This might explain why **2** is only formed in minute quantities, resulting in a partial binding-site occupancy (67%).

To further investigate which of the molecules from the reaction mixture bind to EP, we performed different NMR experiments in the presence of EP. In a time-resolved series of standard 1D ^{1}H-NMR experiments, we were able to corroborate the reactivity of **1** in an EP-containing sample (Figure S7.1B). A 2D ^{1}H-^{1}H-NOESY NMR experiment confirmed the correct structure assignment to **4** and demonstrated that this molecule binds to EP not only in the crystal but also in solution as indicated by positive NOE cross-peaks (Figure 7.2 and Figure S7.2B), which were negative in the absence of EP. An additional ^{1}H-^{13}C-HSQC experiment underlined that this molecule in fact contains an aldehyde function (Figure S7.5). Furthermore, the NOESY spectra confirmed that also **1** can interact with EP (Figure 7.1B and Figure 7.2). Overall, our NMR experiments could identify **1** and **4**, but not **2**, as EP binders. Presumably,

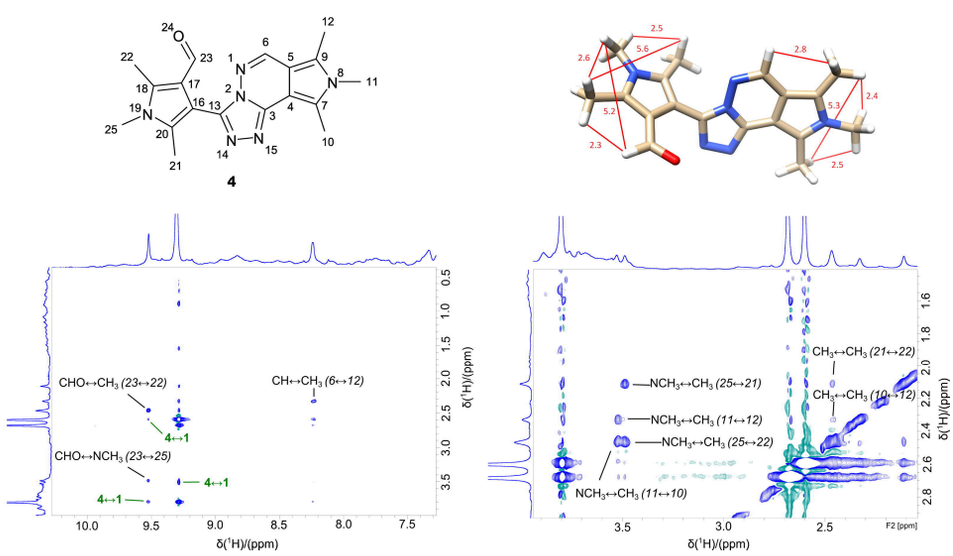


Figure 7.2. NMR analysis of the reaction mixture in the presence of EP. ^{1}H-^{1}H-NOESY NMR spectrum collected from an EP sample incubated with **1** for 12 hours. Positive peaks are shown in blue, negative peaks in cyan. The cross-peaks annotated in black correspond to intramolecular NOE contacts of **4**. Additional intermolecular contacts (green) between **4** and **1** indicate a spatial proximity of both weak binders in the binding pocket of EP. Observed NOE peak intensities match the interatomic distances of the three-dimensional model of **4** (highlighted in red). High-intensity peaks in both spectrum excerpts are signals of the excess **1**.

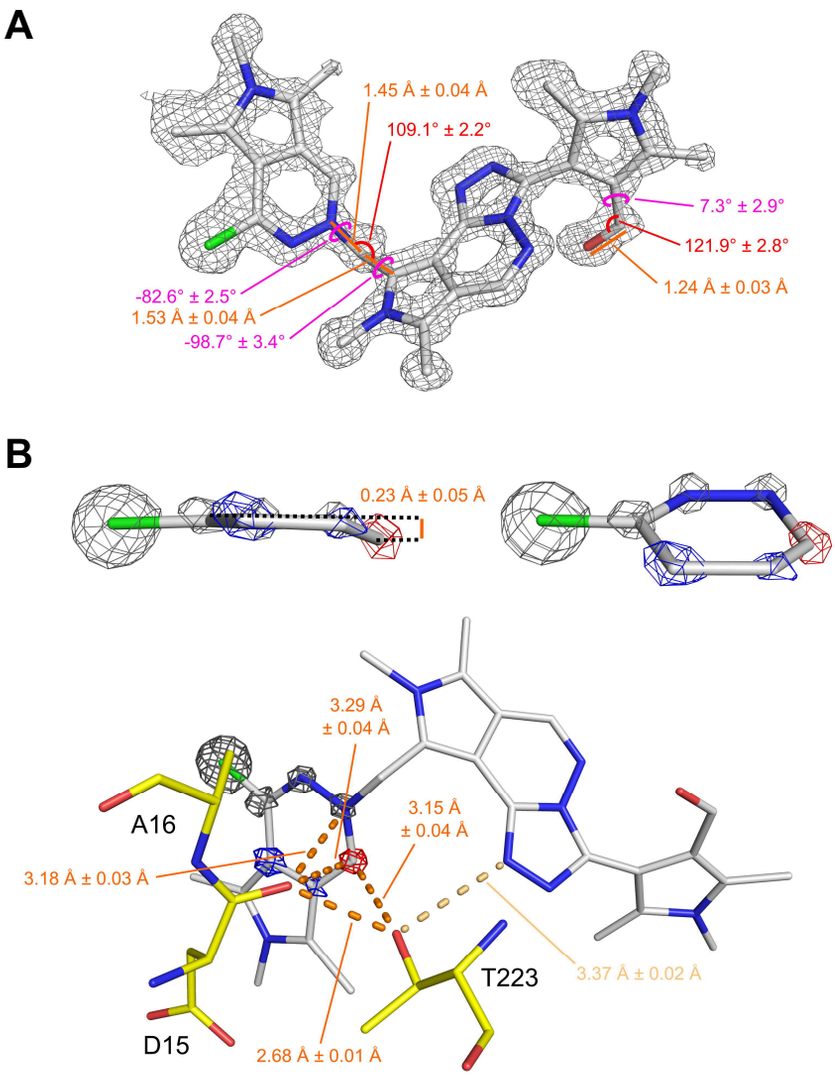


Figure 7.3. Crystallographic elucidation of the unknown structure of **2**. (A) Selected geometric parameters and estimated standard deviations underlining the presence of an aldehyde function and methylene bridge in **2**. The depicted values and $2mF_o$-F_c map (1σ level) were derived using the SHELXL-based refinement strategy described in the Supporting Information and compared to CSD queries (Figure S7.4 and Figure S7.5). (B) Non-planarity of the chloropyridazinium ring. Our SHELXL refinement revealed that one of the pyridazinium carbons is significantly shifted out of the ring plane (4.3 σ). To further highlight the positional accuracy achieved by the refinement of this high-resolution structure, the $2mF_o$-F_c electron density is shown at 1.7 σ for this out-of-plane atom in red as well as for two other pyridazinium carbon atoms in blue. A slightly higher σ-cutoff was chosen for the residual atoms of this positively charged heterocycle (2.3 σ, gray meshes) in order to better visualize the location of each atom as indicated by the density maxima. How the partial positive charge on the slightly pyramidalized carbon atom and attached nitrogen might be compensated by the protein environment is indicated in orange. Based on distance criteria, Thr223 likely points with its hydroxyl hydrogen toward one of the Asp15 carbonyl oxygen lone pairs so that one of the Thr223 lone pairs is directed toward the pyramidalized carbon of **2**. The second Asp15 CO lone pair, in contrast, can compensate the partial positive charge on the nearby pyridazinium nitrogen.

the strong interactions of **2** in the EP binding site, its low level of formation or high reactivity in solution prevented a spectroscopic detection. The binding affinity of **2** can be estimated to be better than 10 μM, as judged by the crystallographic observation of this ligand despite its presence at only μM or even lower concentrations in the soaking solution. The determination of an exact binding constant, however, seems to be hardly possible because **2** is formed in only minute quantities and is very reactive as shown by our QM-calculations, both yet indicating that the ligand might bind even more tightly to EP.

While the finding that **1** and **4** both can weakly interact with EP supported our hypothesis that the unknown EP-binder **2** is composed of these two compounds as suggested by its electron density (Figure 7.1A), it was crucial to show that these two molecular parts are actually covalently linked in the crystal structure. Indeed, the electron density observed in the initial crystal structure could not be explained by the presence of either **1** or **4** in partial occupancy but only by the presence of both parts at the same time in a distance too close to represent separated molecules. Fortunately, the EP-**1**-**2** structure, subsequently obtained from a crystal treated with the HPLC isolate **1**, was of such high resolution (1.03 Å, Figure 7.1B) that it enabled a least-squares refinement using SHELXL and thereby the determination of geometric parameters of the bound molecule **2** along with estimated standard deviations [277]. In particular, we analyzed the geometry of the methylene bridge that must have formed upon connection of **1** and **4** and compared it to values derived from small-molecule crystal structures, clearly indicating that a covalent linkage is chemically consistent and thus very likely (Figure 7.3A and Figure S7.4). Based on our EP-**1**-**2** structure, the presence of an aldehyde in **2** could also be confirmed by its derived geometric characteristics, such as a typical CO bond length of 1.24 Å ± 0.03 Å (Figure 7.3A and Figure S7.5). In summary, this made us confident to assign structure **2** to the initially unknown EP inhibitor. Importantly, an LC/HR-MS experiment indicated the presence of a compound with the proposed molecular composition of **2** in the crude reaction mixture. An MS/MS fragmentation of this isolate, moreover, supported this assignment (Scheme S7.1).

For a hypothetical reaction of **1** and **4** to **2**, the activation of an aliphatic CH_3 group and the subsequent substitution of a hydrogen atom would be a mechanistic requirement. Since 2-methylpyrroles and related 2-methylindoles are known to react at their respective CH_3 groups after irradiation with visible light, anodic or catalytic oxidation [278–281], a radical mechanism was assumed to be involved in the formation of **2**. To investigate the photochemical reactivity of the methyl groups, a sample of the parent compound **1** was irradiated in a photochemical reactor. Compounds **8** and **9** were identified as main products that had been formed under

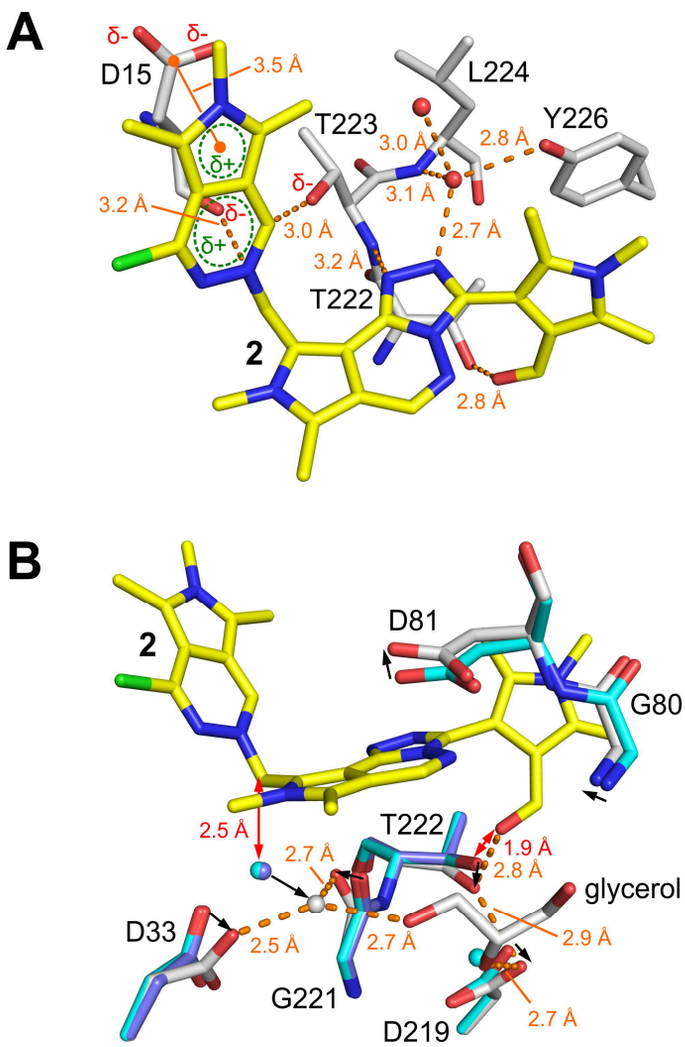


Figure 7.4. Recognition of **2** by EP. (A) Interactions between **2** (yellow) and EP (gray), assessed via intermolecular distances (orange). For more details concerning the proposed electrostatic interactions between Asp15, Thr223 and **2** please see Figure 7.3 and Figure S7.6. (B) Induced-fit binding of **2**. Compared to the EP apo form (cyan, PDB-code 4Y5L), the ligand-bound protein underwent significant conformational changes as indicated by black arrows (RMSD of depicted protein atoms = 0.44 Å). Since the binding site of **2** is not fully populated in the EP-**1**-**2** structure (occupancy of 63%), some of the depicted residues are found in the apo (violet) as well as in the ligand-bound (gray) conformation (e.g. Asp33). Red arrows indicate clashes that would occur between the apo state and the ligand without conformational adaption of the protein.

photochemical activation of **1** in methanol (Scheme 7.1). While this indicated a certain photochemical reactivity, the formation of **2** was also observed in the absence of light. Since **2** was traceable in solution in minute quantities, EP might selectively trap and thereby stabilize this highly reactive molecule. Alternatively, it seems possible that the juxtaposition of **1** and **4**

in the EP binding pocket may directly facilitate the final reaction between these two species upon activation of **4**.

EP specifically recognizes the final reaction product **2** via three hydrogen bonds, one of which is mediated by water (Figure 7.4A). In particular, the aldehyde oxygen of **2** accepts an H-bond from Thr222 although this residue usually acts as an acceptor [66]. This phenomenon is part of several structural adaptations that take place upon binding of the ligand. **2** would sterically interfere with the *apo* protein, which therefore experiences an induced fit as shown in Figure 4B and simultaneously takes up a glycerol molecule from the soaking buffer at the position of the catalytic water. Finally, the question arises how the positive charge of **2** is stabilized within the EP binding pocket. This charge will be partially located within the pyrrole heterocycle that can electrostatically interact with the negatively charged Asp15 stacking on top of the ring (Figure 7.4A). In addition, the least-squares refinement of the EP-**1**-**2** structure revealed a very interesting chemical phenomenon that further clarifies the charge stabilization. The carbon atom attached to the formally positively charged pyridazinium nitrogen is significantly shifted out of the ring plane (4.3 σ, Figure 7.3B). A partial transfer of the negative charge from Thr223 to this imine carbon and the accompanying hybridization change might explain the local breakdown of ring planarity. Since the presence of a positive charge in the ring system of the pyridazine moiety in **2** leads to its destabilization according to our quantum chemical calculations, these electrostatic interactions seem to be key for its stabilization within the binding pocket. In small-molecule crystal structures analogous pyridazinium and pyridinium rings are generally found to be fully planar (Figure S7.6A-B). Two host-guest complexes, however, are exceptions from this trend and bear astonishing similarities to the EP-**1**-**2** protein-ligand complex (Figure S7.6C-E). All of these complexes exhibit two partially negatively charged oxygen atoms in the host next to the positively charged guest molecule. A partial electron transfer from the host to the guest might explain the observed local deviation from planarity.

In summary, we conclude that the presumably weak affinity of **1** to EP alone cannot fully explain the astonishing biological activity observed in multiple assays. Instead, a number of other molecules, which are generated from **1** in a reaction cascade leading to a profound rearrangement of its heterocyclic core structure, contribute to EP inhibition. The unequivocally non-fragment like molecule **2** seems to exhibit the strongest interaction of all reaction products, with an estimated affinity of at least 10 μM if not decidedly much better. It contacts EP via multiple H-bonds and electrostatic interactions that are enabled by the unique chemistry of the inherently cationic compound. This study shows that in addition to a thorough investigation of chemical stability of fragment library members, structural information is essential for the identification of false-positive binders, which is of utmost importance in the early stages of a fragment-based drug discovery project to avoid investments

made to follow up futile leads. Driven by crystallographic data, we were able to identify the 1-chloropyrrolopyridazine core of **1** as a moiety representing an intriguing case of a self-reacting fragment that may be generally prone to react under conditions commonly employed in biological assays.

7.3 Experimental Section

The reaction cascade emanating from **1** was simply initiated by dissolving the compound. Purchased **1** as well as all HPLC isolates have been soaked into EP crystals prepared as described previously before collection of diffraction data and structure determination [64,67]. Additional details about these and all further experiments are provided in the Supporting Information.

7.4 Acknowledgements

We thank the Helmholtz-Zentrum Berlin for the allocation of synchrotron radiation beamtime and travel support. Furthermore, we are grateful for ideas and feedback on the proposed mechanism and refinement results provided by Ali Al-Mourabit, Andrea Vasella and his group, as well as George M. Sheldrick. The work described herein was financially supported by the BMBF-Project Frag2Xtal (No. 05K13RM1) and by the European Research Council (ERC) of the European Union (grant 268145-DrugProfilBind). RL acknowledges the Liebig- (VCI) and Emmy-Noether-Programs (DFG).

7.5 Supporting Information

7.5.1 Supporting Figures

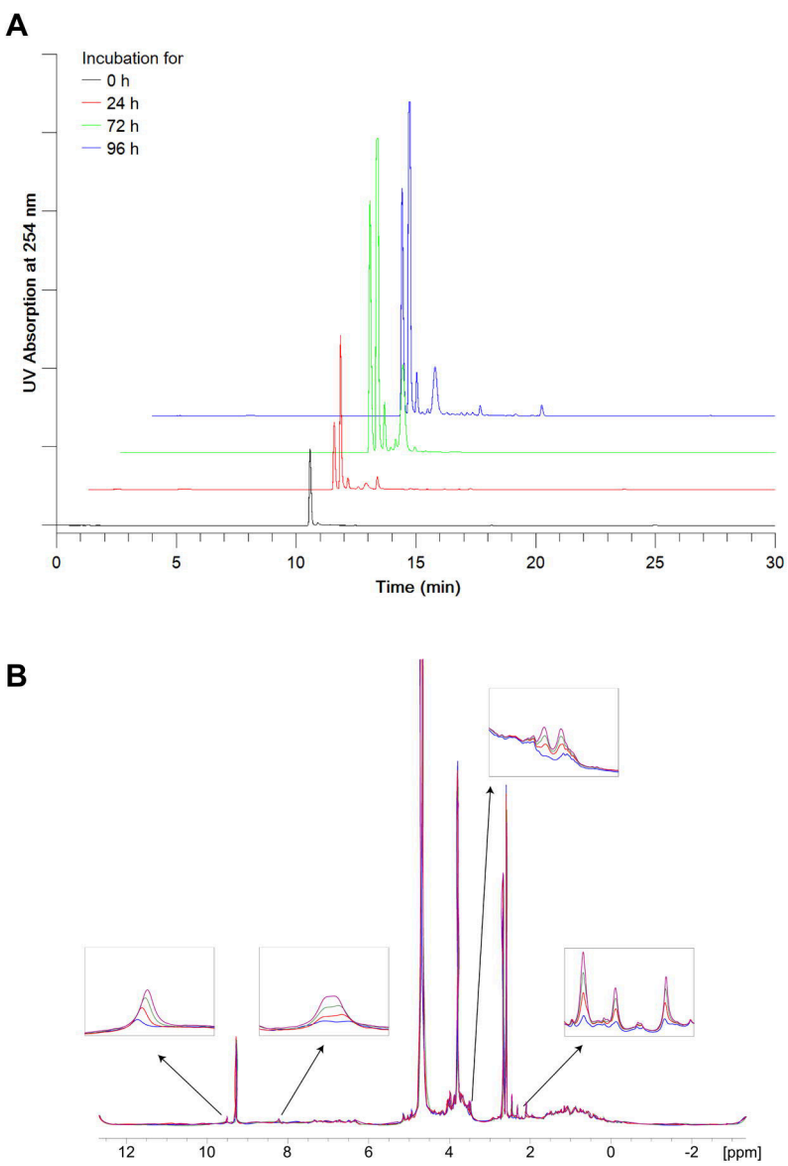


Figure S7.1. Monitoring the reaction of **1** by HPLC and NMR spectroscopy. (A) HPLC chromatogram of a sample of **1** after incubation for 0, 24, 72, and 96 h at 50 °C in methanol. (B) Time-resolved ^{1}H-NMR spectrum of **1** in the presence of EP. Over the course of 12 h, spectra were recorded at 1 h intervals. For clarity, the figure only displays spectra every 3 hours, namely after 0 (blue), 3 (red), 6 (green), and 9 hours of incubation (violet). The major peaks are derived from **1**, whereas additional peaks come up over time.

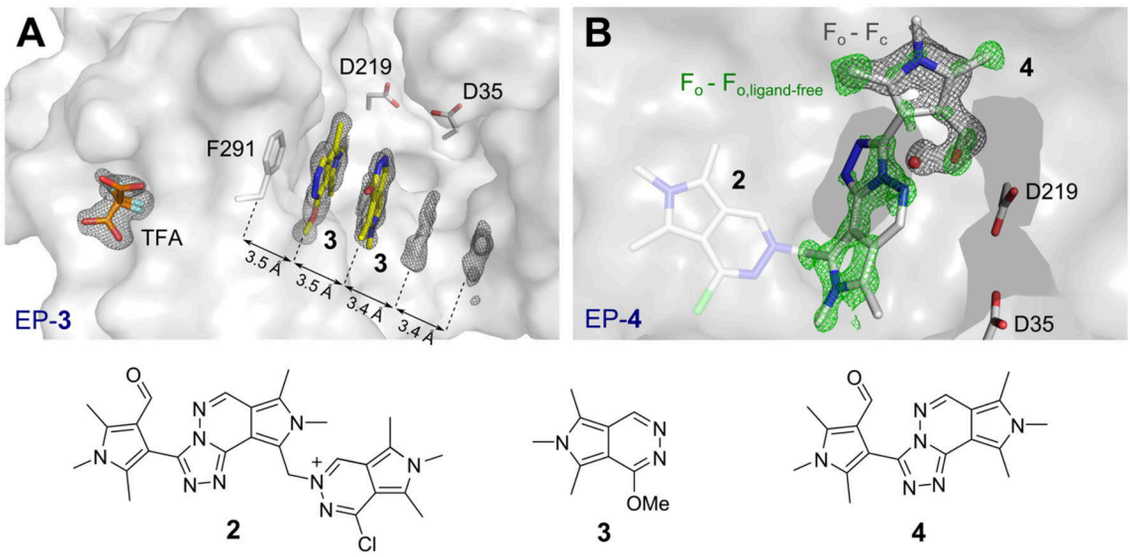


Figure S7.2. EP in complex with reaction products of **1**. The mF_o-DF_c electron density maps for bound ligands are depicted as gray meshes at the 3σ level prior to the inclusion of each respective ligand into the model. (A) Interaction between EP and molecule **3**. The depicted structure EP-**3** is the result from a diffraction experiment with a crystal soaked with the HPLC isolate m/z = 192.3. Electron density has been observed for four copies of compound **3** stacking on top of Phe291 in a regular arrangement. The density was sufficient to model the two molecules closest to Phe291 (the second molecule has only been partially modeled). The TFA molecule shown in two alternative conformations in orange stems from the HPLC purification and binds in a small pocket on the surface of EP. (B) Electron density for molecule **4** in the EP binding pocket after soaking of **5**, the precursor of **4**. While the isolated molecule **4** itself did not bind with sufficient occupancy to the enzyme to make it visible in the electron density, the presence of **4** was indicated by soaking experiments with the better water-soluble compound **5** that is easily converted into **4** (Scheme 7.1 in the main manuscript). For better orientation, the picture contains molecule **2** as a stick model in gray, derived from a superimposition of the EP-**1**-**2** onto the EP-**4** structure (gray surface representation). The bicyclic component of **2** that differentiates it from molecule **4** is depicted in transparent colors. Presumably due to its poor aqueous solubility, **4** does not sufficiently occupy the EP binding site to allow modelling it in the EP-**4** structure. However, for several reasons, we are very confident that **4** binds to EP in a similar manner as the corresponding part of **2**. In particular, the mF_o-DF_c (gray mesh) electron density map indicates the presence of the substituted pyrrole ring of **4**, which is further supported by the observation of the alternative conformations of several residues arising from the induced-fit binding discussed in Figure 7.4B. Presumably, the simultaneous presence of apo and **4**-bound EP molecules in the crystal also results in electron density next to the aldehyde group of **4** originating from a water molecule of the apo protein, which is represented by the red sphere derived from a superimposition of the apo structure (PDB-code 4Y5L) onto the EP-**4** structure. For clarity, the mF_o-DF_c map is only shown around the pyrrole ring. Moreover, an isomorphous difference map (depicted in green) between the EP-**4** and a ligand-free EP structure of an isomorphous crystal (PDB-code 5P71) provides some evidence for the presence of the tricyclic core of **4** (gray sticks). In contrast, no density could be identified for the atoms of the bicyclus additionally present in **2** (transparent gray sticks).

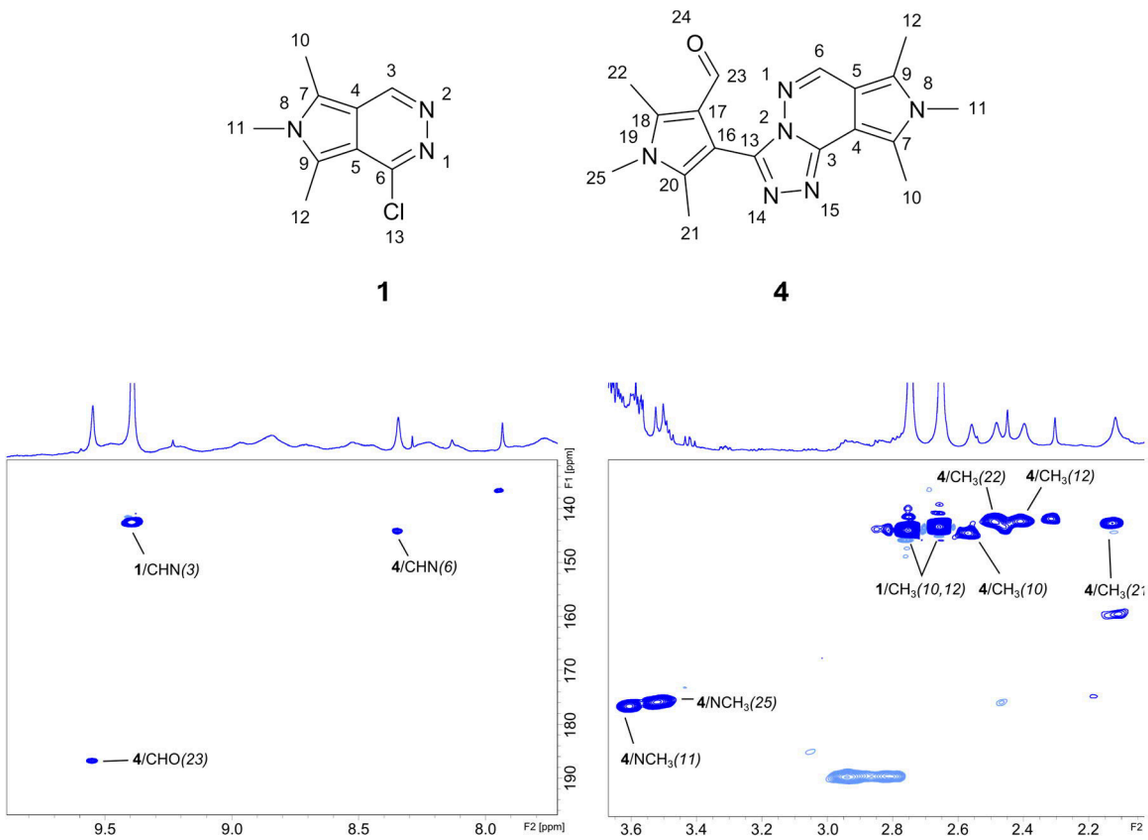


Figure S7.3. ^{1}H-^{13}C-HSQC NMR spectrum collected from an EP sample incubated with **1** after an incubation time of three weeks at 7 °C. Cross-peaks corresponding to **1** and **4** are annotated in the spectrum. Additionally observed signals originate from an unknown molecule, presumably generated from hydrolysis of the 3-chloropyridazine moiety in **1**. An aldehyde function is clearly present in **4** as indicated by a ^{1}H chemical shift of 9.53 ppm and a ^{13}C chemical shift of 186.6 ppm.

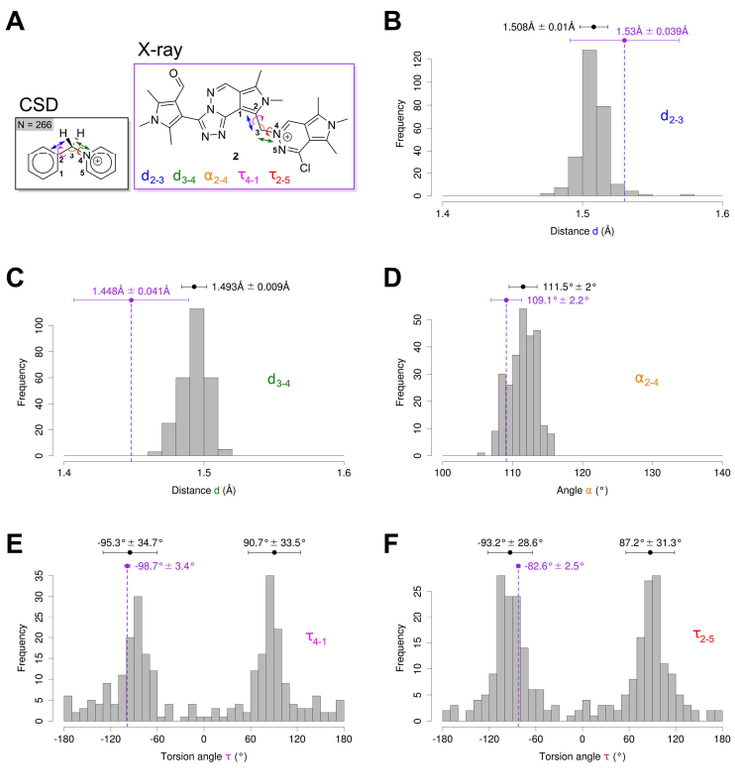


Figure S7.4. Comparison of geometric parameters for the proposed methylene bridge of molecule **2** derived from the crystallographic experiment (EP-**1**-**2** structure) with results from CSD searches for comparable compounds. (A) Definition of the CSD query and investigated geometric parameters. In order to investigate whether molecule **4** and **1** are fused to compound **2** via a methylene bridge, the geometric parameters highlighted in blue, green, orange, magenta and red derived from the crystallographic experiment have been compared with the distributions of the same parameters resulting from a CSD query on comparable molecules as defined in the gray box. In the following panels, the mean values and standard deviations of the individual histogram peaks are given on top of each plot and compared to the crystallographically identified value for **2** including estimated standard deviation (purple, according to the SHELXL refinement performed as described in the experimental section). These analyses reveal that **2** likely harbors an intact methylene bridge. (B) Distribution of the distances between atoms 2 and 3 as defined in panel A. (C) Histogram for the distance between atoms 3 and 4. (D) Distribution of the methylene bridge angles between atoms 2,3 and 4. (E) Histogram for the torsion angle between atoms 4, 3, 2 and 1. If less than 180° rotation of the bond between atoms 4 and 3 in a clockwise fashion (indicated by the magenta arrow in panel A) is required in order to create the eclipsed conformation with the bond between atoms 2 and 1, the torsion angle is considered positive (else negative) according to IUPAC. (F) Histogram for the torsion angle between atoms 2, 3, 4 and 5. If less than 180° rotation of the bond between atoms 2 and 3 in a clockwise fashion (indicated by the red arrow in panel A) is required in order to create the eclipsed conformation with the bond between atoms 4 and 5, the torsion angle is considered positive (else negative). All histograms have been generated using the statistical framework R.

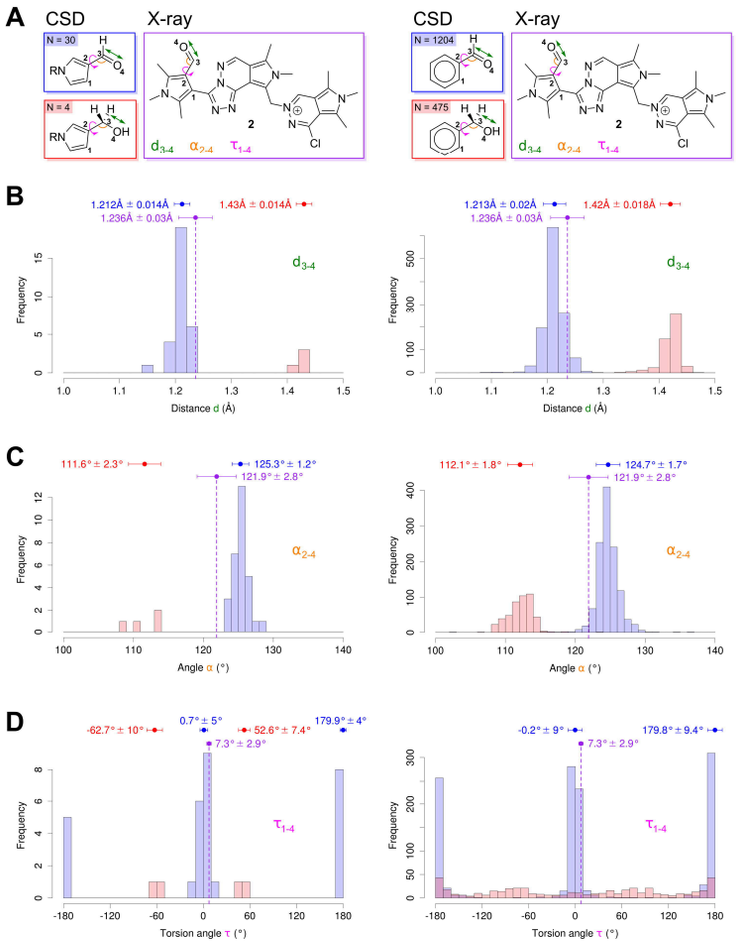


Figure S7.5. Comparison of geometric parameters for the proposed aldehyde group of molecule **2** derived from the crystallographic experiment (EP-**1**-**2** structure) with results from CSD queries for comparable compounds. (A) Definition of the CSD queries and investigated geometric parameters. To decide whether **2** contains an aldehyde (blue boxes in panel A and transparent blue bars in panels B-D) or alcohol (red) functionality, CSD queries were performed with pyrrole-3-carboxaldehydes and pyrrol-3-ylmethanols (left). Since the number of structures containing these motifs in the CSD was rather limited (N = 30 and 4, respectively), similar queries have been performed on much more abundant benzaldehydes and benzylalcohols (N = 1204 and 475, respectively). The three investigated parameters are defined in green, orange and magenta. For instance, the distance between the aldehyde/alcohol carbon and oxygen is highlighted in green. (B) Distribution of aldehyde (blue) vs. alcohol (red) bond lengths (between atoms 3 and 4). The mean values and standard deviations of the individual histogram peaks are given on top of each plot and compared to the crystallographically identified value for **2** including estimated standard deviation (purple, according to the SHELXL refinement performed as described in the experimental section). This comparison clearly indicates that compound **2** contains an aldehyde as supported by similar analyses of the associated angle and dihedral (panels C and D). (C) Histogram for the angle between atoms 2, 3 and 4 (for the definition of these numbers see panel A). (D) Distribution of torsion angle values between atoms 1, 2, 3 and 4. If less than 180° rotation of the bond between atoms 1 and 2 in a clockwise fashion (indicated by the magenta arrow in panel A) is required in order to create the eclipsed conformation with the bond between atoms 3 and 4, the torsion angle is considered positive (else negative) according to IUPAC. These figures have been prepared using the statistical program R.

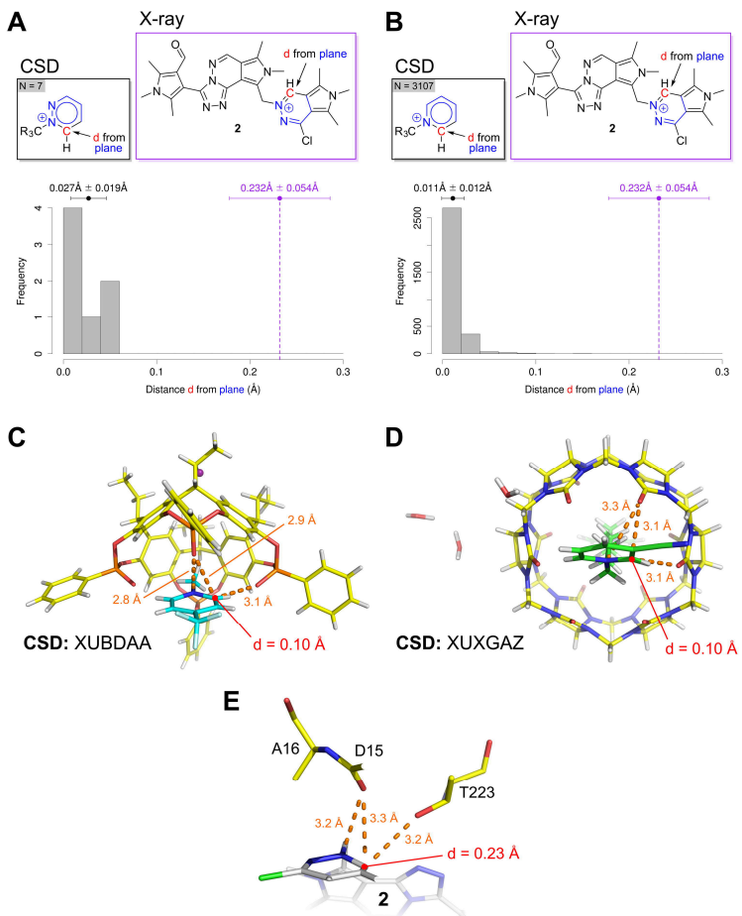


Figure S7.6. Non-planar pyridazinium and pyridinium cations. (A) Comparison of the non-planarity of the pyridazinium heterocycle of molecule **2** as derived from the crystallographic experiment (EP-**1**-**2** structure) with usual geometric features of pyridazinium ions. A CSD query was set up as defined in the gray box. The below histogram reports the distribution of distances between the carbon atom highlighted in red from the plane that is made up by the residual atoms of the heterocycle (out-of-plane deviation). The mean values and standard deviation of the single histogram peak is given on top of the plot and compared to the crystallographically identified value including estimated standard deviation (purple, according to the SHELXL refinement performed as described in the experimental section). (B) Comparison of the non-planarity of the pyridazinium heterocycle of molecule **2** with usual geometric features of pyridinium ions. Since only seven pyridazinium moieties are currently present in the CSD fulfilling our search criteria (for details see also the experimental section), we additionally searched for much more frequent pyridinium ions (N = 3107) and plotted the results in the same way as described for panel A. Both analyses underline that these heterocyclic cations are usually planar. However, some exceptions with significantly shifted carbon positions out of the ring plane do exist. We visually analyzed all small molecule crystal structures

with d > 0.08 Å including their packing and found that this deviation can frequently be attributed to unusually twisted and strained molecular structures as found e.g. in helicenes. Two outliers, however, are very interesting with respect to the observed phenomenon (panels C and D). Both are host-guest complexes that bear similarities to the EP-**1-2** protein-ligand complex (panel E). All complexes have two partially negatively charged oxygen atoms in close vicinity to the positively charged pyridinium/pyridazinium nitrogen and attached slightly pyramidalized carbon atom. The latter atom is highlighted in red along with the distance from the plane while important intermolecular distances are shown in orange. (C) A host-guest complex containing a significantly non-planar pyridinium ion (CSD-code XUBDAA) [300]. (D) A rotaxane containing a slightly distorted pyridinium ion (CSD-code XUXGAZ) [301]. (E) The high-resolution EP-**1-2** crystal structure containing a non-planar pyridazinium moiety. The environment of the pyridazinium cation is highlighted. All histograms have been generated using the statistical framework R while structural figures have been prepared using Pymol.a clockwise fashion (indicated by the magenta arrow in panel A) is required in order to create the eclipsed conformation with the bond between atoms 3 and 4, the torsion angle is considered positive (else negative) according to IUPAC. These figures have been prepared using the statistical program R.

7.5.2 **Supporting Tables**

Table S7.1. X-ray diffraction data collection and refinement statistics.

	Complex (PDB code)			
	EP-**2** (5LWR)	EP-**1**-**2** (5LWS)	EP-**3** (5LWT)	EP-**4** (5LWU)
(A) Data collection and processing				
Wavelength (Å)	0.91841	0.91841	0.91841	0.91841
Beamline	BESSY BL14.2	BESSY BL14.1	BESSY BL14.1	BESSY BL14.1
Detector	MARMOSAIC 225	PILATUS 6M	PILATUS 6M	PILATUS 6M
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters				
a, *b*, *c* (Å)	45.2, 72.8, 52.6	45.3, 73.1, 53.0	45.4, 73.1, 52.7	45.3, 73.3, 53.1
α, β, γ (°)	90.0, 109.2, 90.0	90.0, 109.8, 90.0	90.0, 109.6, 90.0	90.0, 109.9, 90.0
(B) Diffraction data				
Resolution range (Å)	42.7-1.25 (1.32-1.25)	41.2-1.03 (1.09-1.03)	42.7-1.07 (1.13-1.07)	42.6-1.11 (1.18-1.11)
Unique reflections	87290 (12864)	151655 (23423)	142071 (22805)	128260 (20551)
$R(I)_{sym}$ (%)	3.8 (31.8)	5.5 (51.5)	5.1 (48.5)	4.7 (48.6)
Wilson *B* factor (Å^2)	10.5	9.3	9.0	10.3
Completeness (%)	98.1 (89.7)	94.6 (90.6)	99.5 (98.9)	99.6 (99.0)
Redundancy	4.0 (3.0)	3.9 (3.8)	3.7 (3.6)	3.7 (3.6)
$\langle I/\sigma(I)\rangle$	20.7 (3.5)	11.8 (2.3)	13.1 (2.3)	13.5 (2.3)
(C) Refinement				
Resolution range (Å)	28.1 - 1.25	36.5 - 1.03	29.4 - 1.07	39.8 - 1.11
Reflections used in refinement (work/free)	82903 / 4363	144059 / 7581	134943 / 7103	121835 / 6413
Final *R* value for all reflections (work/free) (%)	11.3 / 13.1	11.6 / 13.2	12.8 / 14.8	12.3 / 14.3
Protein residues	330	330	330	330
Ligand atoms	38	49	26	
Other ligand atoms	44	47	43	42
Water molecules	329	350	312	326
RMSD from ideality				
Bond lengths (Å)	0.006	0.007	0.007	0.006
Bond angles (°)	0.9	1.0	1.0	1.0
Ramachandran plot[c]				
Most favored regions (%)	93.9	93.5	94.2	94.2
Additionally allowed regions (%)	6.1	6.5	5.8	5.8
Generously allowed regions (%)	0	0	0	0
Disallowed regions (%)[d]	0	0	0	0
Mean B factor (Å^2)				
Protein non-hydrogen atoms	12.0	11.5	11.6	12.6
Main Chain	11.0	10.5	10.7	11.7
Side Chain	12.8	12.5	12.5	13.4
Ligand	11.7	17.1	30.7	
Other ligand atoms	21.6	23.2	25.3	25.8
Water molecules	29.1	26.1	24.8	27.7

[a] Values in parenthesis refer to the highest resolution shell [b] *Other ligands are glycerol, ethylene glycol, DMSO, acetate and TFA* [c] Calculated using PROCHECK [138] [d] Calculated using MOLEMAN [182].

7.5.3 Experimental Section

General Experimental Information

1-Chloro-5,6,7-trimethyl-6H-pyrrolo[3,4-d]pyridazine (1) was purchased as a hydrochloride salt from Enamine (UKR). ^{1}H and ^{13}C spectra were recorded on a JEOL ECX-400 or Bruker AV II 300 instrument. All chemical shift values are reported in ppm relative to the non-deuterated solvent signal. Trimethylsilyl propanoic acid was used as an external standard for ^{13}C NMR spectra in D_2O. To describe the multiplicity of the signal, the following abbreviations were used: s = singlet, q = quartet, m = multiplet. ESI-MS spectra were recorded on a Q-Trap 2000 system by Applied Biosystems. For high resolution ESI-MS, MS/MS and LC/MS a LTQ-FT Ultra (Thermo Fischer Scientific) or Orbitrap Velos Pro (Thermo Fisher Scientific) mass spectrometer were used. For HPLC chromatography a Shimadzu LC-20 system equipped with a diode array detector was used. Analytic separations were carried out with a MN Nucleodur 100-5 C18ec 4.6 × 250 mm column using a water-acetonitrile gradient with the addition of 0.1% TFA. For semi-preparative separations a Waters XSelect CSH C18 10 × 250 mm column using a water-acetonitrile gradient was used. Preparative separations were carried out on a Varian PrepStar 218 instrument equipped with a MNagel Nucleodur 100-5 C18ec 32 x 250 mm column, employing a water-acetonitrile gradient with the addition of 0.1% TFA.

Ageing of Compound 1

To initiate the reaction cascade, a sample of **1** was placed in a clear glass flask or GC-vial, dissolved in methanol (Fisher, HPLC grade) and stirred at rt or 50 °C. Preliminary experiments were carried out in EP protein buffer (100 mM NaOAc pH 4.6), in the presence and absence of EP, respectively, as well as in water at various conditions. Since these experiments did not give any additional insight and are comparable to the results in methanol, the details are not reported here. For analytical HPLC and MS experiments, 1–5 mg of **1** were dissolved and the resulting solution was used as is. The experiment was also carried out with exclusion of sunlight (brown glass vial, rt, 24 h) and under protective gas (Argon, rt, 24 h).

For a semi-preparative separation 20 mg of **1** (HCl salt, 0.09 mmol) were used, the solvent methanol was removed after the reaction (24 hours, 50 °C) under reduced pressure and the resulting residue was taken up in 1 mL MeOH/H_2O 1:9. The crude product was purified on a semi-preparative HPLC system equipped with a Waters XSelect CSH C18 (10 x 250 mm) column using a water/acetonitrile gradient. After lyophilization, compound **4** (4 mg, 0.01 mmol, 13%) was isolated as a light brown solid as the main product of the reaction. 12 mg (0.05 mmol, 57 %) of the starting material could be recovered.

The preparative separation was carried out with 100 mg of **1** (HCl salt, 0.43 mmol). The solvent methanol was removed after the reaction (24 hours, 50 °C) under reduced pressure and the resulting residue was taken up in 2 mL MeOH/H_2O 1:9. The crude product was purified on a preparative HPLC system using a water/acetonitrile gradient with the addition of 0.1% TFA. Fractions that contained a reaction product were collected and lyophilized. Compound 3 (TFA salt, < 1 mg, < 3.29 µmol, < 0.76%) was isolated as a light brown solid. Compound 5 (TFA salt, 31 mg, 0.07 mmol, 15%) was isolated as a red oil as the main product of the reaction. Compound 6 (TFA salt, < 1 mg, < 1.55 mmol, < 0.4%) was isolated as a red oil. 38 mg (0.12 mmol, 29 %) of the starting material could be recovered.

Photochemical Activation of 1

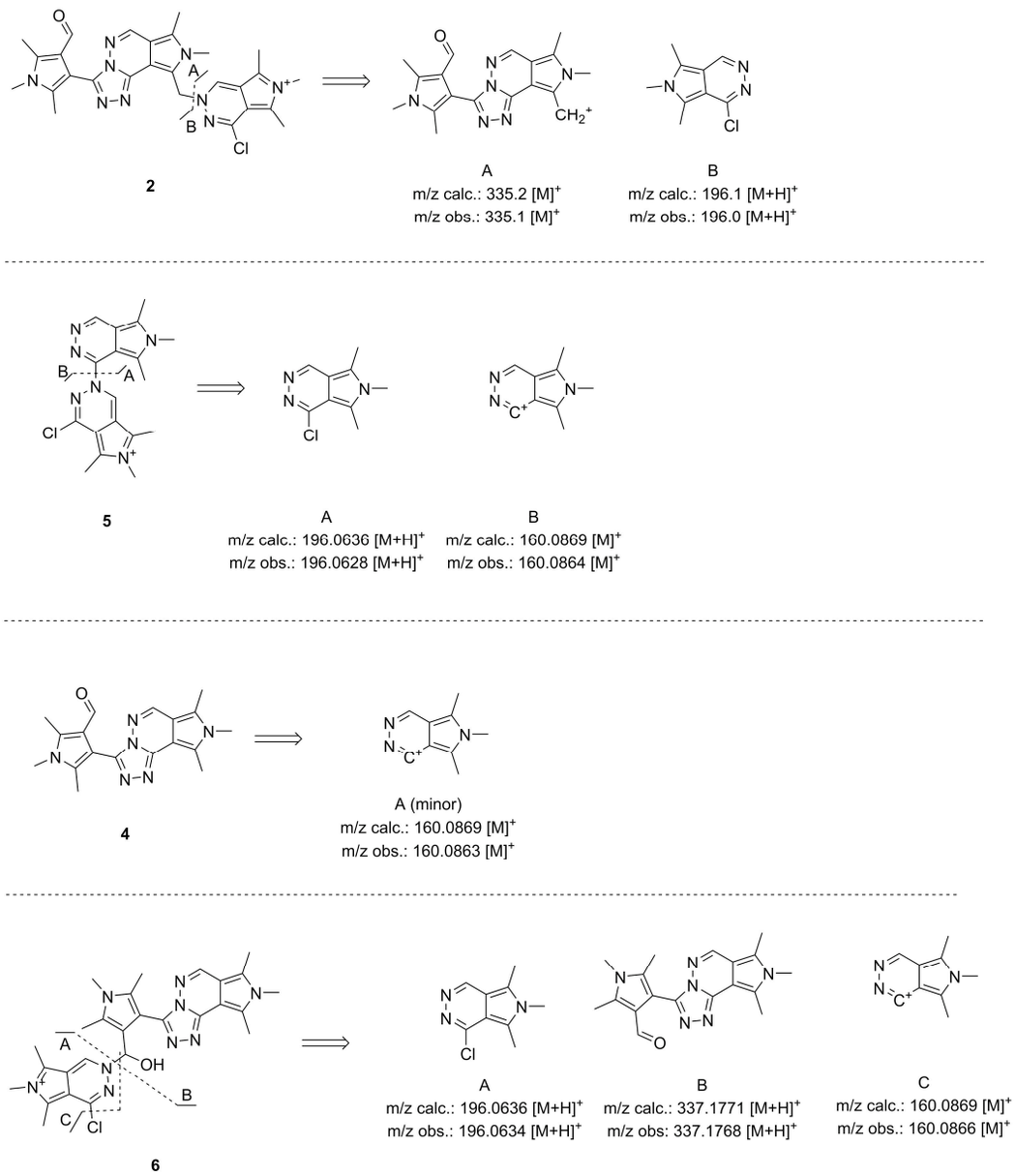
A sample of **1** (HCl salt, 108 mg, 0.47 mmol) was dissolved in 200 mL of methanol and placed in an immersion-type photochemical reaction apparatus equipped with a 150 W medium pressure Hg lamp (Hanau TQ 150). The solution was irradiated for 150 min and the solvent was subsequently removed under reduced pressure. The crude product was taken up in 2 mL MeOH/H_2O 1:9 and subjected to a preparative HPLC separation. After lyophilization, the TFA salt of compound **9** (26 mg, 0.07 mmol, 15%) was isolated as a dark oil as the main product of the reaction.

Conversion of Compound 5 to Compound 4

A sample of **5** was placed in a screw cap vial and treated with 10% ammonia solution. After 24 h the sample was analyzed by HR-ESI-MS.

Tandem MS analysis of Compounds 2, 5, 4 and 6

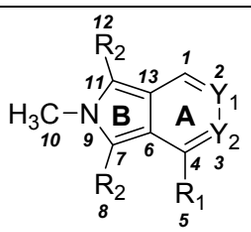
MS/MS experiments of compounds **2**, **5, 4** and **6** were conducted on an Orbitrap Velos Pro (Thermo Fischer Scientific) mass spectrometer using an unpurified reaction mixture (24 hours, 50 °C). Because of a superimposed signal, compound **2** could not be isolated in the ion trap. Instead it was analyzed by LC/MS/MS on a LTQ-FT Ultra (Thermo Fischer Scientific) instrument. Due to the low concentration of **2**, an ITMS2 configuration had to be used, which resulted in a reduced resolution of the spectra.



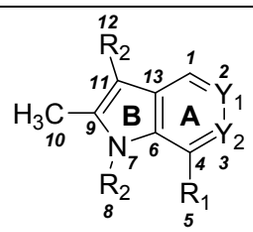
Scheme S7.1. Tandem MS fragmentation patterns observed for **2**, **5**, **4** and **6**. Experimentally observed HRMS data is compared to calculated m/z values of the postulated fragments.

Quantum Chemical Calculations

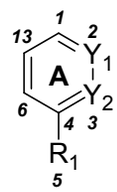
Table S7.2. Overview of compounds investigated by quantum chemical calculations. Compounds I-X and P-X refer to bicyclic structures (left) while compounds B-X and Pyr-X refer to monocyclic structures (right).



isoindole ([3',4']-pyrrolo)



indole ([2',3']-pyrrolo)



monocyclic

Compound	Scaffold	Y_1	Y_2	R_1	R_2
I-1	isoindole	C	C	H	H
I-2	isoindole	C	C	H	CH_3
I-3	isoindole	C	C	Cl	CH_3
I-4	isoindole	*C*-CH_3	C	Cl	CH_3
P-1	isoindole	N	N	H	H
P-2	isoindole	N	N	H	CH_3
P-3 (1)	isoindole	N	N	Cl	CH_3
P-4	isoindole	N^+-CH_3	N	Cl	CH_3
P-5	indole	N	N	H	CH_3
P-6	indole	N	N	Cl	CH_3
B-1	monocyclic	C	C	H	-
B-2	monocyclic	C	C	Cl	-
Pyr-1	monocyclic	N	N	H	-
Pyr-2	monocyclic	N	N	Cl	-

Structures were initially build with MOE [151] and optimized at the B3LYP/6-31+G(d) level using Gaussian09 [282]. No imaginary frequencies were found at the stationary points, indicating that the optimization converged to a true energy minimum. The optimized geometries were used as input structures for single point SCF calculations using different basis sets (6-31G(d), 6-31+G(d) and 6-311+G(d)). The data trends discussed in the main text refer to calculations on the 6-311G(d) level of theory. The detailed results are described in Tables S5.3 to S5.16.

NICS calculation: Absolute NMR-shielding tensors were calculated using the GIAO method [283]. The in-plane ring centers (NICS(0)) and out-of-plane ring centers (NICS(1.0)) were calculated based on the positions of carbon and nitrogen atoms only.

NBO Analysis: Wiberg bond indices (WBI) and natural atomic charges were calculated using NBO 3.1[284] as implemented in Gaussian09.

Electrophilicity index ω: The electrophilicity index ω was calculated according to *Parr et al.* [285] as $\omega = \mu^2/(2\eta)$. The chemical potential μ is defined as $\mu = 0.5 \cdot (E_{LUMO} + E_{HOMO})$ and the chemical hardness η is defined as $\eta = 0.5 \cdot (E_{LUMO} - E_{HOMO})$.

Example Gaussian input file:

%chk=I-1_6-311+Gd

%nproc=2

#B3LYP/6-311+G* SCF=tight Test freq nmr pop=nboread IOp(10/46=1) gfprint

remark line goes here

<<Coordinates>>

$NBO RESONANCE NPA NBO NBOSUM BNDIDX E2PERT NLMO DIPOLE NRT PLOT $END

NOESY and HSQC Experiments in the Presence of Endothiapepsin

NMR measurements were carried out with 300 μl of protein/ligand solution (148 μM EP, 13 mM compound **1**) in a 50 mM CD_3CO_2Na buffer at a pD of 5.1 in a Shigemi tube. The time-resolved ^{1}H- and the [^{1}H-^{1}H]-NOESY spectra (τ_m=2 s) were measured on a Bruker 800 MHz ^{1}H frequency spectrometer, equipped with a triple-resonance cryoprobe, at 298K. 12 consecutive ^{1}H- and 5 consecutive [^{1}H-^{1}H]-NOESY spectra were recorded every hour, starting from mixing the protein and the ligand. The [^{1}H-^{1}H]-NOESY spectra were recorded with 96152(t_2)*272(t_1) complex points, t_{1max}=14.2 ms, t_{2max}=3.0 s, 4 scans, interscan delay 0.3 s. The HSQC spectra were measured three weeks after mixing the protein and the ligand, at a Bruker 600 MHz ^{1}H frequency spectrometer, equipped with a triple-resonance cryoprobe, at 298K. The HSQC experiment was recorded as two separate experiments, focusing on the aliphatic and on the aromatic and carbonyl ^{13}C chemical shift range, respectively, each with 120 ppm spectral width in the indirect dimension, to ensure complete excitation of all ^{13}C nuclei. The phase-sensitive, echo/anti-echo-edited HSQC spectra were recorded with 1024(t_2)*128(t_1) complex points, t_{1max}=3.5 ms, t_{2max}=53.2 ms, 192 scans, and interscan delay of 2 s.

Crystallization, Soaking and X-ray Diffraction Data Collection

EP was extracted from Suparen samples, kindly supplied by DSM Food Specialties, as described previously [286]. Subsequently, the protein was crystallized in its apo-form upon streak-seeding using the vapor diffusion method with a mother liquor composed of 100 mM NH_4OAc, 100 mM NaOAc pH 4.6, 24-30% (w/v) PEG 4000 at 17 °C [64,67] The originally purchased compound **1** has been soaked into EP crystals once at a concentration of 45 mM

and thrice at 90 mM resulting in the structures EP-**1**-**2** and EP-**2**, respectively. Following a similar approach, all soaking experiments with the compounds isolated via HPLC (**1**, **3**, **4**, **5** and **6**) have been performed at two different concentrations of 90 and 250 mM. At 45 and 90 mM, soaking was performed for 48 h at 17 °C in 70 mM NH_4OAc, 70 mM NaOAc pH 4.6, 16-20% PEG 4000, 23% glycerol, 9% DMSO and 45 or 90 mM of the ligand, respectively. At the higher 90 mM ligand concentration, EP crystals have been transferred to this solution via a 1:1 mixture of this solution with the crystallization mother liquor. The same procedure was also applied at 250 mM where the soaking solution consisted of 65 mM NH_4OAc, 65 mM NaOAc pH 4.6, 14-18% PEG 4000, 10% glycerol, 25% DMSO and 250 mM ligand. In contrast to the 90 mM soaks, the duration of the experiment had been reduced to 24 h under the more harsh conditions at 250 mM. Finally, all crystals were directly flash-frozen in liquid nitrogen prior to data collection at the BESSY MX beamlines BL14.1 and BL14.2 [224]. Subsequently diffraction data have been processed using XDS [185].

Structure Determination

To be able to evaluate and compare the outcome of all diffraction experiments thoroughly, we used our automated refinement pipeline [68] to determine structural models for all collected data sets via molecular replacement using Phaser [159] and several Phenix-based refinement steps [162]. Following this approach, the data set with the highest quality and clearest ligand electron density has been chosen for each ligand-bound structure (EP-**2**: 45 mM, EP-**1**-**2**: 90 mM, EP-**4**: 90 mM, EP-**3**: 250 mM). Subsequently, the associated models were manually further refined against structure factor amplitudes until convergence. Library files for all non-standard ligands were generated using the Grade web server application [287]. For ligand **2**, the obtained restraints have been modified in a way to allow the pyridazinium bridge head nitrogen and attached carbon to deviate from the ring plane as observed in the unrestrained SHELXL refinement described below. In order to obtain a reliable estimate of the real binding-site occupancy of **2**, we first refined the occupancy of the system Asp33-H_2O-Gly221-Thr222 that appeared in two alternative conformations, one of which represents the apo-form and one the ligand-bound conformation (Figure 7.4B). After validation of the refined value via visual inspection of the anisotropic displacement parameters, the resulting occupancy for the latter conformational state was assigned also to ligand **2** (67% for EP-**2** structure and 63% for EP-**1**-**2**). Finally, all resulting structures and structure factor amplitudes have been deposited in the PDB. Data collection and refinement statistics are given in Table S7.1. Figures displaying structural information have been prepared using Pymol [164].

Estimation of Standard Deviations for Different Geometric Parameters

To enable an estimation of standard deviations for geometric parameters, a least-squares BLOC refinement was performed against intensities using SHELXL [277]. As a starting point, the final Phenix-refined EP-**1**-**2** structure has been chosen and adapted in several modeling and refinement steps. For the subsequent determination of distances and angles within ligand **2**, all of its atoms were refined in an unrestrained manner except for planarity restraints for the three ring systems excluding the carbon atom next to the pyridazinium nitrogen which was observed to be significantly out-of-plane (Figure 7.3B). This observation was made via a similar refinement in which the planarity restraint for the bicyclic component of **2** was additionally omitted and the deviation of each atom of the pyridazinium ring from a least-squares plane made up by the residual five atoms of the ring was calculated along with estimated standard deviations. While the bridge head pyridazinium nitrogen was found to be located significantly below the plane (-0.20 ± 0.05 Å, corresponding to -4.1 σ) and the attached carbon atom accordingly above the plane (0.23 ± 0.05 Å, corresponding to 4.3 σ), all other atoms did not deviate as significantly from their respective planes (-0.11 to 0.06 Å, -2.1 to 1.1 σ). In order to ensure that these results had not been biased by the ligand restraints used in the previous Phenix-based refinement steps, we repeated both analyses with starting structures that contained an alcohol CO bond length of 1.41 Å instead of the aldehyde bond length of 1.22 Å and an approximately planar pyridazinium ring, respectively. Importantly, the results did not deviate significantly from those obtained before.

Queries in the Cambridge Structural Database (CSD)

CSD searches have been performed using Conquest [288] vs. 1.18 based on CSD vs. 5.37 as of November 2015 (with 2 updates) while the results were analyzed with Mercury [289] vs. 3.8 and plotted via the statistical program R [290]. The following adjustments have been used for all searches: 3D coordinates determined, only organics, R ≤ 0.075, not disordered, no errors and not polymeric. The investigated chemical structures and geometric parameters are defined in Figure S7.4, Figure S7.5 and Figure S7.6.

7.5.4 Experimental Data

*Compound **1***. **^{1}H NMR** (400 MHz, D_2O) δ = 2.78 (s, 3H), 2.85 (s, 3H), 3.97 (s, 3H), 9.53 (s, 1H). **^{13}C NMR** (101 MHz, D_2O) δ = 10.3, 10.7, 32.8, 110.4, 113.5, 129.6, 136.5, 142.4, 150.9. **HRMS (ESI+)** calculated for $C_9H_{11}ClN_3$: 196.0636 $[M+H]^+$; found: 196.0636.

*Compound **2***. **HRMS (ESI+)** calculated for $C_{27}H_{29}ClN_9O$: 530.2178 $[M]^+$; found: 530.2176.

*Compound **3** (TFA salt).* **^{1}H NMR** (400 MHz, D_2O) δ = 2.63 (s, 3H), 2.65 (s, 3H), 3.81 (s, 3H), 4.11 (s, 3H), 9.15 (s, 1H). **HRMS (ESI+)** calculated for $C_{10}H_{14}N_3O$: 192.1131 $[M+H]^+$; found: 192.1128.

*Compound **4**.*[a] **^{1}H NMR** (400 MHz, MeOH-*d*4) δ = 2.26 (s, 3H), 2.62 (s, 3H), 2.63 (s, 3H), 2.80 (s, 3H), 3.63 (s, 3H), 3.81 (s, 3H), 8.52 (s, 1H), 9.70 (s, 1H). **^{13}C NMR** (101 MHz, MeOH-*d*4) δ = 10.4, 10.8, 11.0, 11.8, 30.9, 31.7, 111.8, 125.3, 133.6, 143.8, 170.3. **HRMS (ESI+)** calculated for $C_{18}H_{21}N_6O$: 337.1771 $[M+H]^+$; found: 337.1767.

*Compound **5** (TFA salt).* **^{1}H NMR** (400 MHz, D_2O) δ = 2.45 (s, 3H), 2.80 (s, 3H), 2.83 (s, 3H), 2.87 (s, 2H), 3.95 (s, 3H), 3.97 (s, 2H), 9.70 (s, 1H). **^{13}C NMR** (101 MHz, D_2O) δ = 10.6, 10.8, 10.9, 12.3, 33.4, 105.6, 109.4, 115.6, 116.2 (q, $^1J_{CF}$ = 288.7), 116.4, 128.5, 132.1, 138.4, 142.9, 144.0, 144.9, 150.4, 151.4, 162.9 (q, $^2J_{CF}$ = 35.2). **HRMS (ESI+)** calculated for $C_{18}H_{20}ClN_6$: 355.1438 $[M]^+$; found: 355.1433.

*Compound **6** (TFA salt).* **^{1}H NMR** (400 MHz, D_2O) δ = 2.32 (s, 3H), 2.54 (s, 3H), 2.75 (s, 3H), 2.79 (s, 3H), 2.87 (s, 3H), 2.93 (s, 3H), 3.89 (s, 3H), 3.95 (s, 3H), 4.04 (s, 3H), 9.74 (s, 1H), 10.15 (s, 1H), 10.44 (s, 1H). **HRMS (ESI+)** calculated for $C_{27}H_{31}ClN_9O$: 532.2335 $[M]^+$; found: 532.2334.

*Compound **8**.* **HRMS (ESI+)** calculated for $C_{10}H_{13}ClN_3O$: 226.0742 $[M+H]^+$; found: 226.0730.

*Compound **9** (TFA salt).* **^{1}H NMR** (400 MHz, D_2O) δ = 3.44 – 3.49 (m, 6H), 4.14 (s, 3H), 5.07 (s, 2H), 5.18 (s, 2H), 9.84 (s, 1H). **^{13}C NMR** (101 MHz, D_2O) δ = 33.8, 57.6, 58.2, 61.5, 63.2, 114.2, 115.6, 116.2 (q, $^1J_{CF}$ = 291.7), 128.0, 135.1, 143.1, 150.9, 162.8 (q, $^2J_{CF}$ = 43.1). **HRMS (ESI+)** calculated for $C_{11}H_{15}ClN_3O_2$: 256.0847 $[M+H]^+$; found: 256.0833.

[a] Due to its poor solubility, compound **4** was dissolved in a mixture of MeOH-d4 and acetone-d6. The low concentration prohibited the observation of several signals of quaternary carbon atoms in the ^{13}C spectrum.

7.5.5 Data from Quantum Chemical Calculations

Table S7.3. Natural atomic charges at atom position 4.

Compound	6-31G(d)	6-31+G(d)	6-311+G(d)
I-1	-0.21638	-0.21759	-0.18359
I-2	-0.21592	-0.21616	-0.18410
I-3	-0.01476	-0.01326	-0.00804
I-4	-0.00645	-0.00394	0.01021
P-1	0.03557	0.03206	0.06757
P-2	0.03546	0.03085	0.08457
P-3 (1)	0.22077	0.20894	0.22538

Table S7.3. continued.

Compound	6-31G(d)	6-31+G(d)	6-311+G(d)
P-4	0.23685	0.22813	0.24438
P-5	-0.03548	-0.03769	-0.00296
P-6	0.15427	0.14564	0.16039
B-1	-0.23527	-0.24218	-0.20333
B-2	-0.23848	-0.24463	-0.20632
Pyr-1	-0.01251	-0.02041	0.02052
Pyr-2	-0.01417	-0.02107	0.01868

Table S7.4. Natural atomic charges at atom position 1.

Compound	6-31G(d)	6-31+G(d)	6-311+G(d)
I-1	-0.21638	-0.21759	-0.18359
I-2	-0.21594	-0.21959	-0.17120
I-3	-0.21468	-0.21542	-0.17048
I-4	-0.21542	-0.21410	-0.18760
P-1	0.03439	0.03198	0.06752
P-2	0.03546	0.03378	0.06459
P-3 (1)	0.03818	0.03664	0.07056
P-4	0.12658	0.13490	0.16506
P-5	0.01105	0.00766	0.04347
P-6	0.01253	0.00908	0.04383
B-1	-0.23527	-0.24218	-0.20333
B-2	-0.04358	-0.05584	-0.04102
Pyr-1	-0.01251	-0.02041	0.02052
Pyr-2	0.15588	0.14349	0.16034

Table S7.5. WBI for atomic bond at position 2-3.

Compound	6-31G(d)	6-31+G(d)	6-311+G(d)
I-1	1.2657	1.2663	1.2669
I-2	1.2604	1.2603	1.2167
I-3	1.2556	1.2542	1.2535
I-4	1.2204	1.2198	1.2167
P-1	1.2191	1.2227	1.2239
P-2	1.2107	1.2129	1.2147
P-3 (1)	1.1936	1.1945	1.1935
P-4	1.1269	1.1302	1.1264
P-5	1.335	1.3374	1.3397
P-6	1.3086	1.3095	1.3097
B-1	1.4368	1.4371	1.439
B-2	1.432	1.4324	1.4327
Pyr-1	1.3976	1.4027	1.4047
Pyr-2	1.3736	1.3765	1.3763

Table S7.6. WBI for atomic bond at position 3-4.

Compound	6-31G(d)	6-31+G(d)	6-311+G(d)
I-1	1.5988	1.5980	1.6005
I-2	1.6013	1.6004	1.5679
I-3	1.5624	1.5598	1.5623
I-4	1.5758	1.5735	1.5679
P-1	1.6445	1.6473	1.6468
P-2	1.6492	1.6504	1.6436
P-3 (1)	1.6370	1.6387	1.6356
P-4	1.5705	1.5672	1.5648
P-5	1.5291	1.531	1.5299
P-6	1.5408	1.5451	1.5437
B-1	1.4368	1.4371	1.439
B-2	1.4037	1.4027	1.4025
Pyr-1	1.4824	1.4866	1.4856
Pyr-2	1.4748	1.4782	1.4745

Table S7.7. NICS(0) values at ring *A*.

Compound	6-31G(d) [ppm]	6-31+G(d) [ppm]	6-311+G(d) [ppm]
I-1	-8.3573	-6.7175	-6.5780
I-2	-7.8801	-6.3974	-6.1650
I-3	-8.1807	-7.3796	-7.4937
I-4	-8.6765	-7.3959	-7.0781
P-1	-5.0951	-4.4809	-4.5637
P-2	-4.6257	-4.1097	-4.1701
P-3 (1)	-5.6378	-4.9806	-4.9024
P-4	-6.2992	-5.5065	-5.2127
P-5	-8.9700	-8.3265	-8.3780
P-6	-8.9700	-8.3265	-8.3780
B-1	-9.6482	-8.0216	-7.9345
B-2	-10.2524	-8.9027	-9.9924
Pyr-1	-5.7556	-5.0531	-5.2485
Pyr-2	-6.4445	-5.8008	-5.8481

Table S7.8. NICS(0) values at ring *B*.

Compound	6-31G(d) [ppm]	6-31+G(d) [ppm]	6-311+G(d) [ppm]
I-1	-17.4253	-15.6495	-15.3332
I-2	-16.8376	-15.3714	-15.2396
I-3	-17.2753	-16.0619	-15.8239
I-4	-16.9905	-15.5229	-15.2236
P-1	-17.9539	-16.4588	-16.0083
P-2	-17.3109	-15.8880	-15.6712
P-3 (1)	-17.4879	-16.4384	-16.1498
P-4	-16.6222	-15.4282	-15.1428
P-5	-13.5300	-12.7020	-12.3425
P-6	-13.9308	-13.1218	-12.6420

Table S7.9. NICSzz(0) values at ring *A*.

Compound	6-31G(d) [ppm]	6-31+G(d) [ppm]	6-311+G(d) [ppm]
I-1	-5.6777	-5.4037	-6.0251
I-2	-3.1337	-2.8154	-3.4881
I-3	-29.4937	-29.6613	-30.4412
I-4	-2.5119	-1.9326	-2.4267
P-1	-2.4149	-2.4149	-3.5926
P-2	0.1742	0.1742	-0.9561
P-3 (1)	0.0075	0.0075	-0.7008
P-4	4.2965	4.6270	4.4654
P-5	-11.4265	-11.9404	-12.6243
P-6	-10.2845	-10.4693	-11.0410
B-1	-13.7989	-13.1765	-14.0643
B-2	-13.3587	-12.6903	-13.3565
Pyr-1	-10.4678	-10.7302	-11.5122
Pyr-2	-9.8905	-9.8762	-10.4523

Table S7.10. NICSzz(0) values at ring *B*.

Compound	6-31G(d) [ppm]	6-31+G(d) [ppm]	6-311+G(d) [ppm]
I-1	-16.6309	-16.1241	-16.6410
I-2	-13.8302	-13.8624	-14.5943
I-3	-13.8886	-13.9385	-14.6399
I-4	-12.6404	-12.2745	-12.9272
P-1	-16.7125	-16.7125	-16.8537
P-2	-13.7984	-13.7984	-14.5550
P-3 (1)	-13.5122	-13.5122	-14.1549
P-4	-12.9942	-12.6259	-13.3833
P-5	-28.3137	-28.2621	-28.3954
P-6	-25.8995	-25.6857	-25.7820

Table S7.11. NICS(1.0) values at ring *A*.

Compound	6-31G(d) [ppm]	6-31+G(d) [ppm]	6-311+G(d) [ppm]
I-1	-9.4985	-8.2738	-8.4818
I-2	-9.0535	-7.9922	-8.1357
I-3	-9.1508	-8.0746	-8.0797
I-4	-8.9200	-7.8523	-7.8267
P-1	-9.2444	-8.3725	-8.5583
P-2	-8.5047	-7.7288	-7.9358
P-3 (1)	-8.2936	-7.4413	-7.6149
P-4	-7.2098	-6.4063	-6.4560
P-5	-11.7000	-10.8731	-11.0804
P-6	-11.1418	-10.3154	-10.5151
B-1	-11.1929	-10.1364	-10.1345
B-2	-11.0419	-9.9924	-10.0023
Pyr-1	-11.0397	-10.2596	-10.4739
Pyr-2	-10.6968	-9.9392	-10.0860

Table S7.12. NICS(1.0) values at ring *B*.

Compound	6-31G(d) [ppm]	6-31+G(d) [ppm]	6-311+G(d) [ppm]
I-1	-13.8887	-12.2812	-12.3669
I-2	-13.6878	-12.2126	-12.4800
I-3	-13.7288	-12.3437	-12.3577
I-4	-13.4765	-12.0651	-12.1919
P-1	-14.1433	-12.7693	-12.8363
P-2	-13.8699	-12.5762	-12.9047
P-3 (1)	-13.7434	-12.5977	-12.6247
P-4	-14.0630	-12.8504	-12.9812
P-5	-11.3494	-10.4291	-10.5042
P-6	-11.4074	-10.4793	-10.4817

Table S7.13. HOMO Energies

Compound	6-31G(d) [eV]	6-31+G(d) [eV]	6-311+G(d) [eV]
I-1	-4.72988	-5.05424	-5.09887
I-2	-4.37777	-4.65396	-4.69968
I-3	-4.64471	-4.8942	-4.93941
I-4	-4.58593	-4.82294	-4.86839
P-1	-5.78840	-6.11440	-6.15957
P-2	-5.33942	-5.61997	-5.66514
P-3 (1)	-5.62486	-5.88065	-5.92555
P-4	-9.84780	-9.98250	-10.02440
P-5	-5.7751	-6.1315	-6.1645
P-6	-6.1476	-6.3922	-6.4390
B-1	-6.6940	-6.9939	-7.0385
B-2	-6.7016	-6.9457	-6.9822
Pyr-1	-6.3465	-6.6997	-6.7332
Pyr-2	-6.8390	-7.1514	-7.1835

Table S7.14. LUMO Energies

Compound	6-31G(d) [eV]	6-31+G(d) [eV]	6-311+G(d) [eV]
I-1	-0.38612	-0.78994	-0.78994
I-2	-0.27946	-0.63456	-0.63456
I-3	-0.59592	-0.91375	-0.91375
I-4	-0.51211	-0.81253	-0.81253
P-1	-1.1428	-1.50479	-1.50479
P-2	-0.97117	-1.28764	-1.28764
P-3 (1)	-1.25417	-1.53227	-1.53227
P-4	-5.92909	-6.07902	-6.07902
P-5	-0.6664	-1.0275	-1.0917
P-6	-0.9285	-1.2251	-1.2912
B-1	0.0841	-0.3951	-0.4620
B-2	-0.3499	-0.7761	-0.8444
Pyr-1	-1.3910	-1.7979	-1.8618
Pyr-2	-1.8153	-2.1742	-2.2376

Table S7.15. HOMO-LUMO Energy Gaps.

Compound	6-31G(d) [eV]	6-31+G(d) [eV]	6-311+G(d) [eV]
I-1	-4.34375	-4.26429	-4.30892
I-2	-4.09830	-4.01939	-4.06511
I-3	-4.04878	-3.98048	-4.02565
I-4	-4.07381	-4.01041	-4.05585
P-1	-4.64553	-4.60961	-4.65478
P-2	-4.36824	-4.33232	-4.37749
P-3 (1)	-4.37069	-4.34838	-4.39328
P-4	-3.91871	-3.90347	-3.94538
P-5	-5.1087	-5.1040	-5.0727
P-6	-5.2191	-5.1672	-5.1479
B-1	-6.7781	-6.5988	-6.5765
B-2	-6.3517	-6.1696	-6.1378
Pyr-1	-4.9555	-4.9019	-4.8714
Pyr-2	-5.0238	-4.9772	-4.9459

Table S7.16. Electrophilicity Index ω.

Compound	6-31G(d) [eV]	6-31+G(d) [eV]	6-311+G(d) [eV]
I-1	1.5064	2.0023	2.0120
I-2	1.3231	1.7396	1.7499
I-3	1.6958	2.1186	2.1276
I-4	1.5949	1.9798	1.9893
P-1	2.5853	3.1484	3.1550
P-2	2.2791	2.7534	2.7608
P-3 (1)	2.7067	3.1593	3.1650
P-4	15.8796	16.5220	16.4319
P-5	2.0305	2.5104	2.5949
P-6	2.3984	2.8073	2.9020
B-1	1.6115	2.0685	2.1386
B-2	1.9571	2.4161	2.4950
Pyr-1	3.0204	3.6827	3.7912
Pyr-2	3.7271	4.3682	4.4864

7.5.6 **Data from NMR, MS and HPLC Experiments**

^{1}H (400 MHz, D_2O) and ^{13}C NMR (101 MHz, D_2O) of compound 1

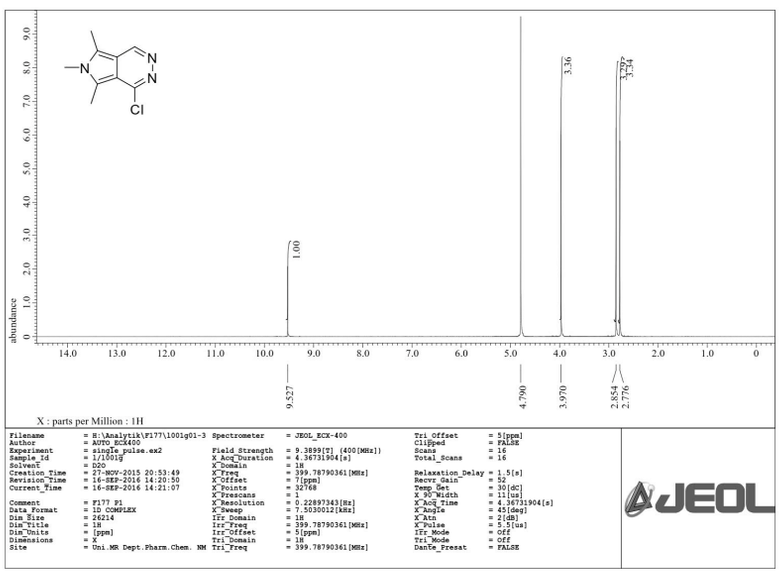


Figure S7.7. ^{1}H NMR Spectrum of **1**.

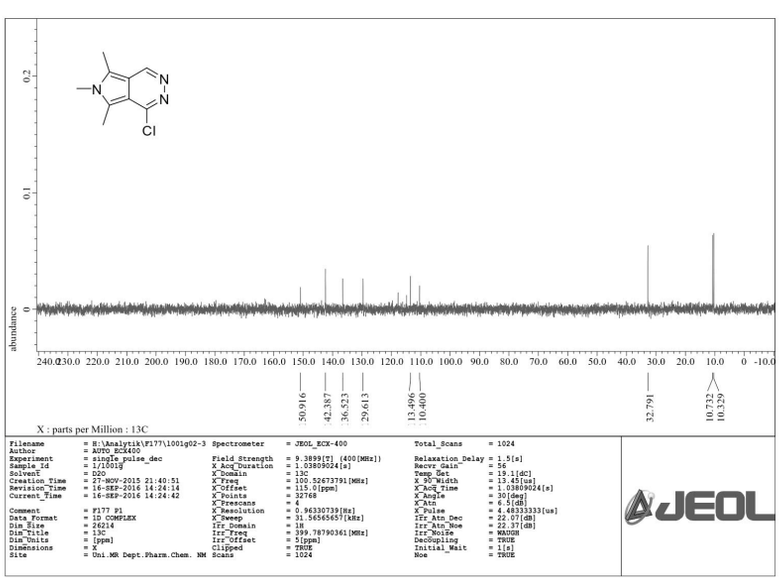


Figure S7.8. ^{13}C NMR Spectrum of **1**.

^{1}H (400 MHz, D_2O) NMR spectrum of compound 3

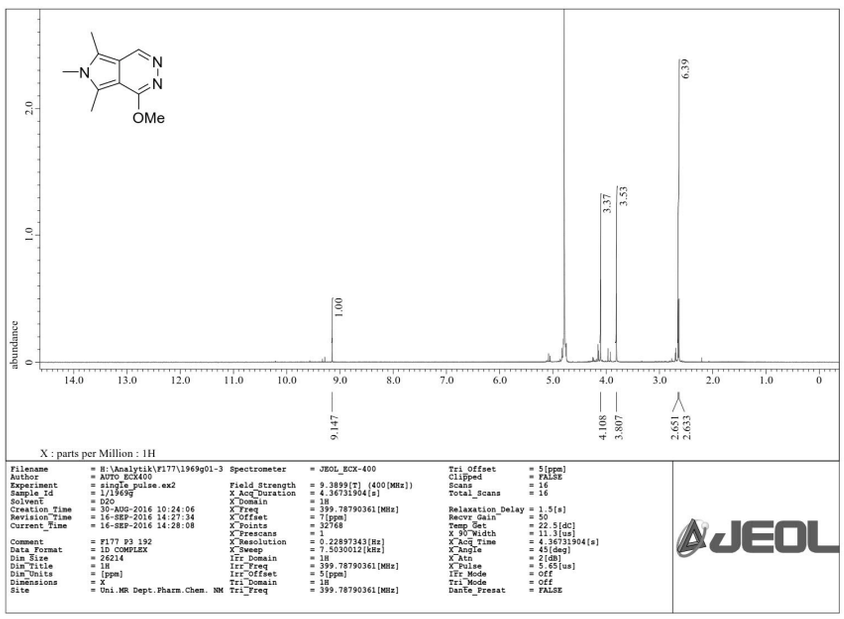


Figure S7.9. ^{1}H NMR Spectrum of **3**.

^{1}H (400 MHz, MeOH-d4) and ^{13}C NMR (101 MHz, MeOH-d4) spectra of compound 4

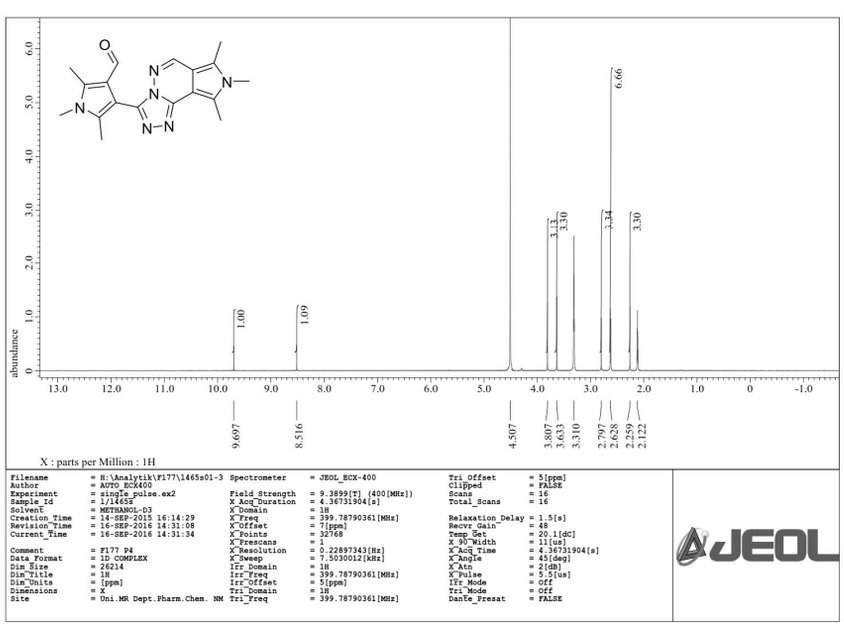


Figure S7.10. ^{1}H NMR Spectrum of **4**.

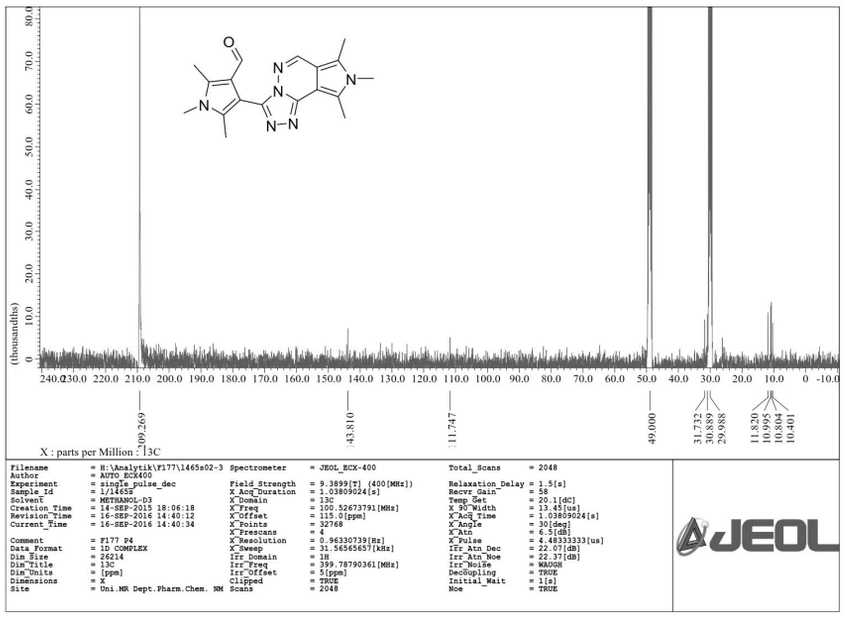


Figure S7.11. ^{13}C NMR Spectrum of **4**.

^{1}H (400 MHz, D_2O) and ^{13}C NMR (101 MHz, D_2O) spectra of compound 5

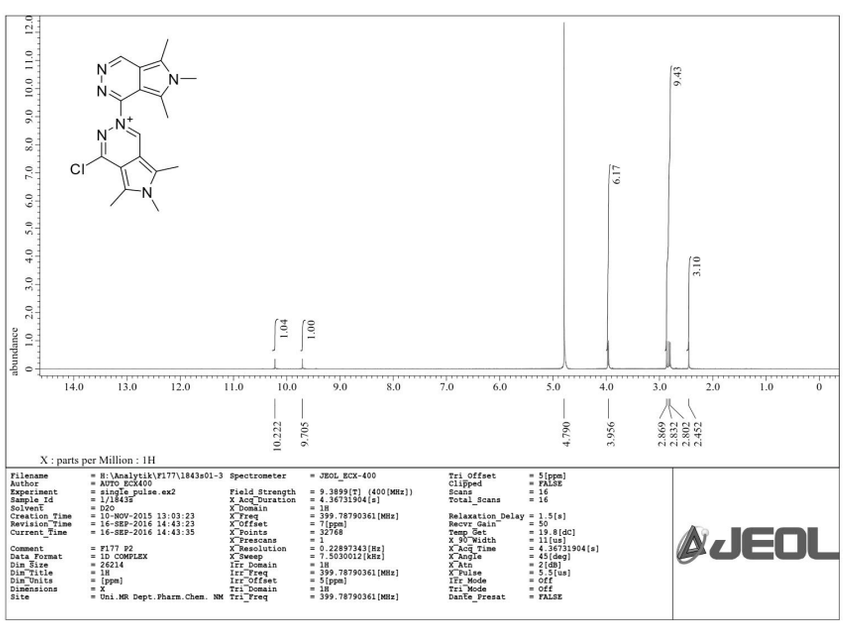


Figure S7.12. ^{1}H NMR Spectrum of **5**.

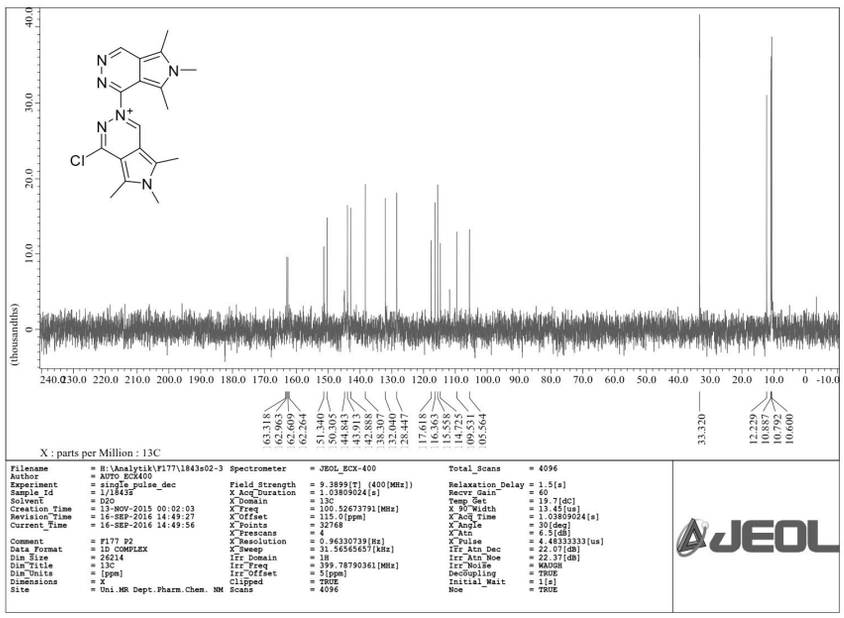


Figure S7.13. ^{13}C NMR Spectrum of **5**.

^{1}H (400 MHz, D_2O) NMR spectrum of compound 6

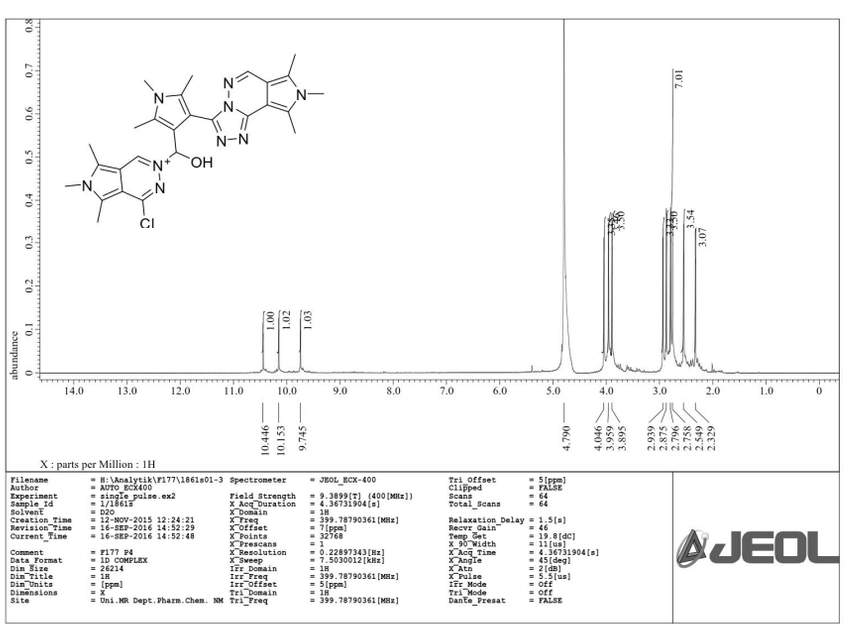


Figure S7.14. ^{1}H NMR Spectrum of **6**.

^{1}H (400 MHz, D_2O) NMR spectrum of compound 9

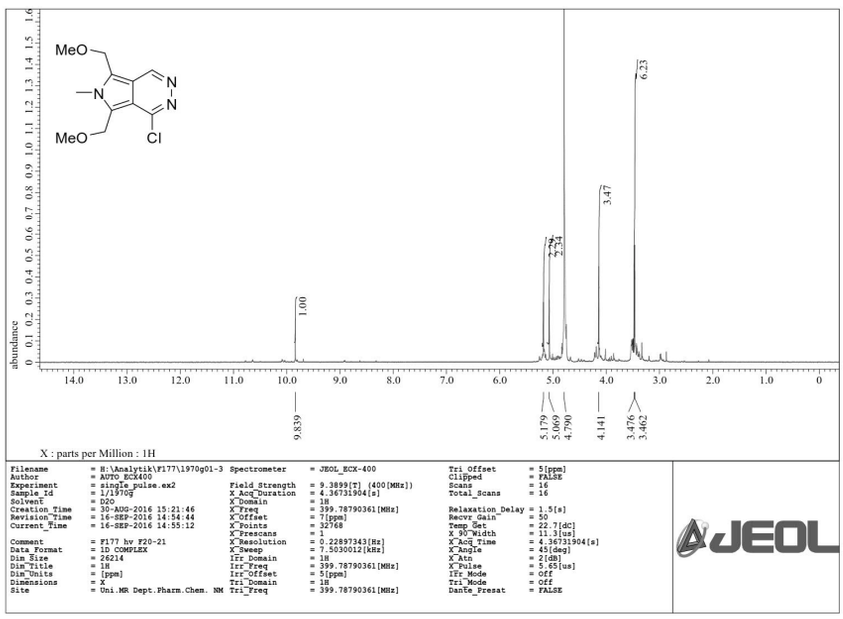


Figure S7.15. ^{1}H NMR Spectrum of **9**.

ESI-MS and HR-MS spectra of compound 1

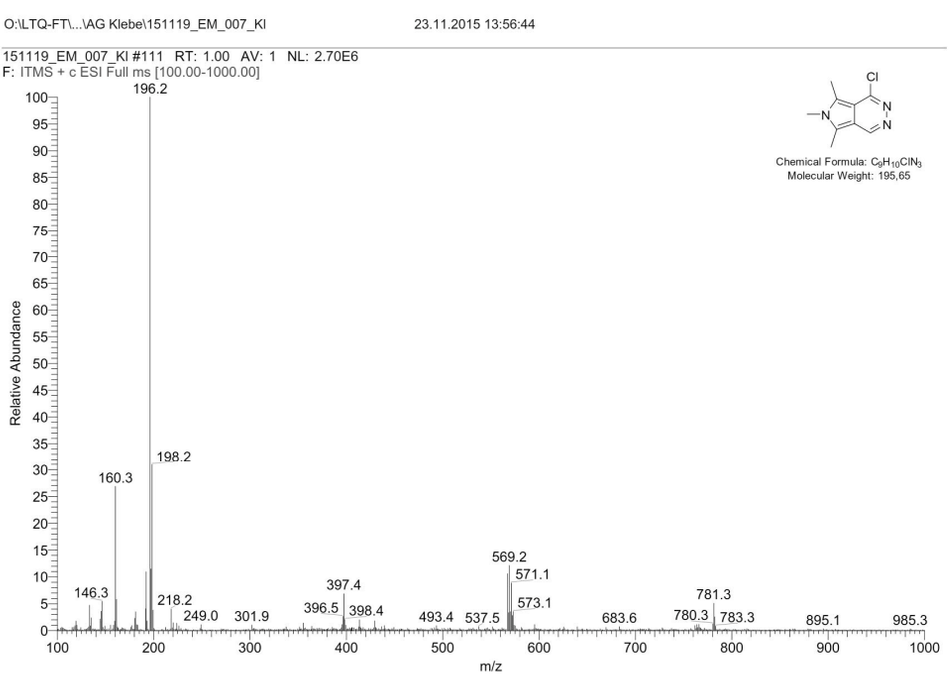


Figure S7.16. ESI-MS Spectrum of **1**.

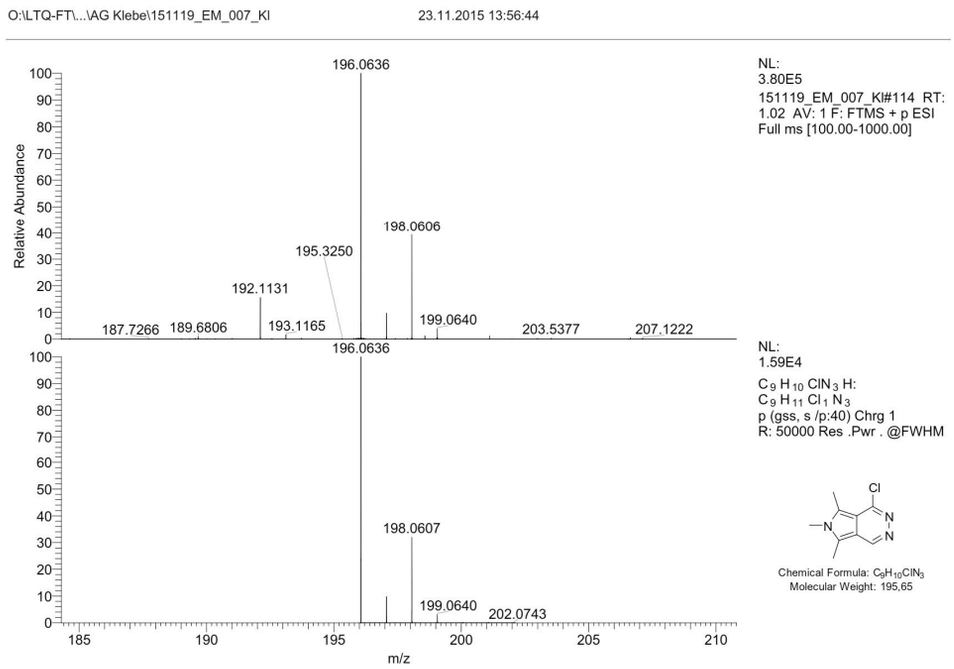


Figure S7.17. ESI-HRMS Spectrum of **1**.

ESI-MS and HR-MS spectra of compound 2

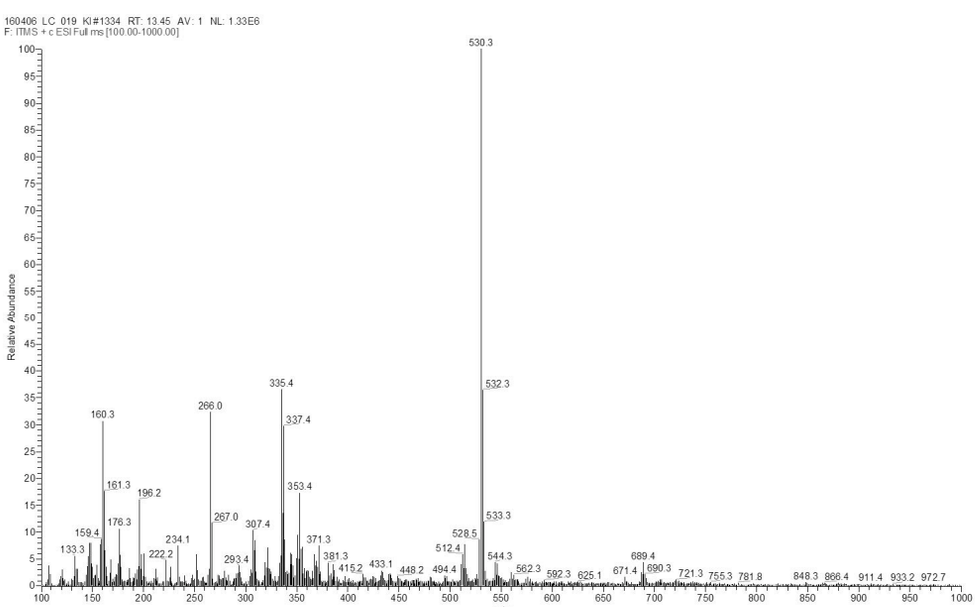


Figure S7.18. ESI-MS Spectrum of **2**.

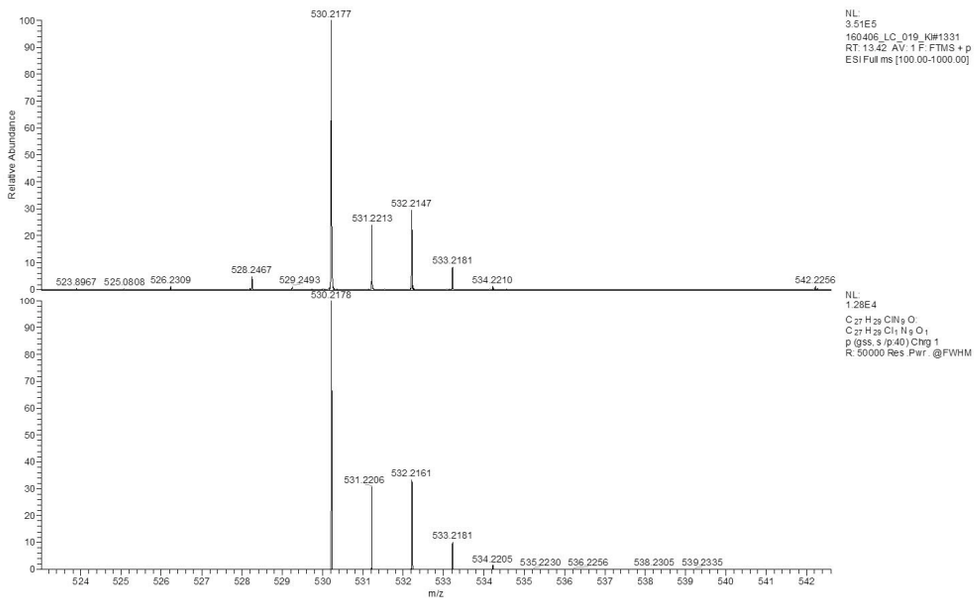


Figure S7.19. ESI-HRMS Spectrum of **2**.

ESI-MS and HR-MS spectra of compound 3

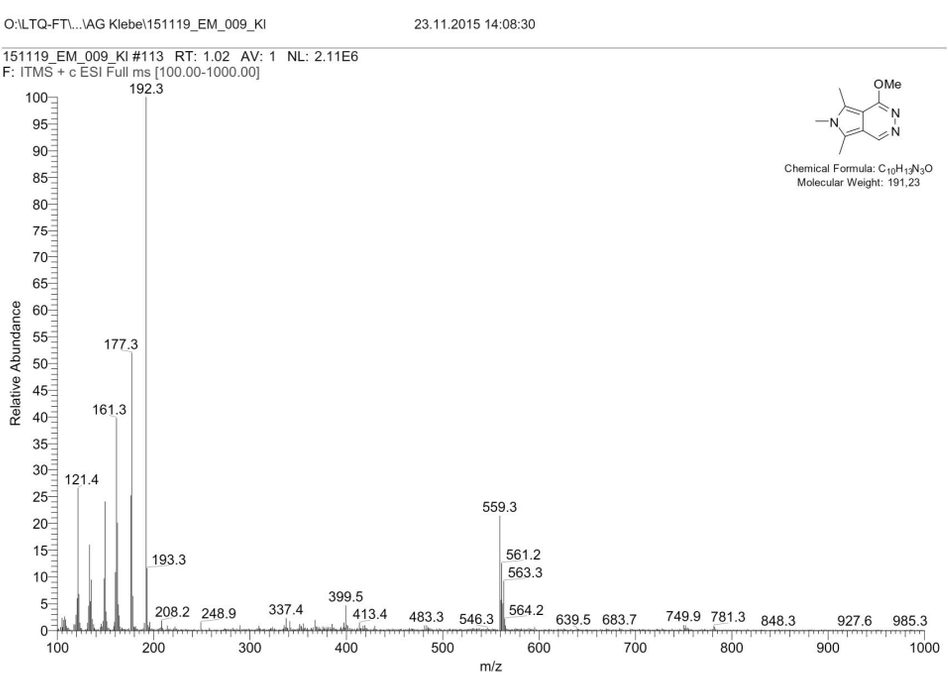


Figure S7.20. ESI-MS Spectrum of **3**.

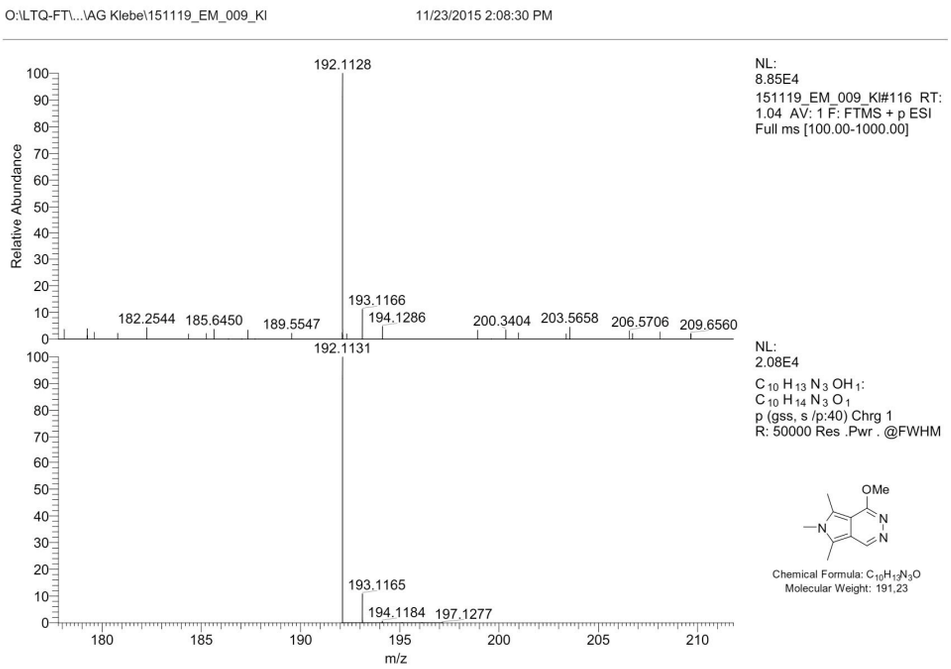


Figure S7.21. ESI-HRMS Spectrum of **3**.

ESI-MS and HR-MS spectra of compound 4

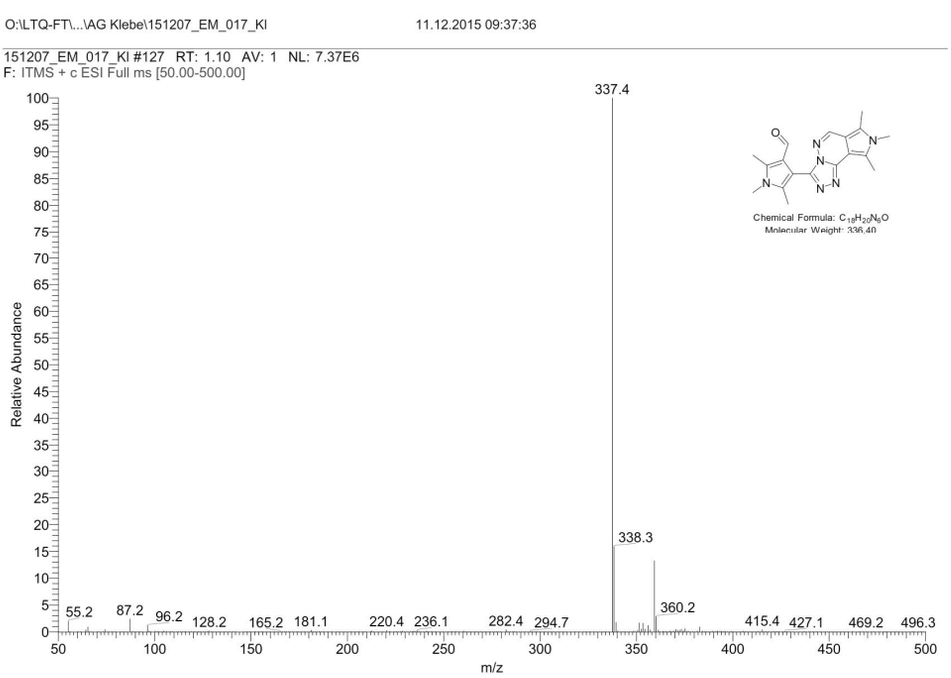


Figure S7.22. ESI-MS Spectrum of **3**.

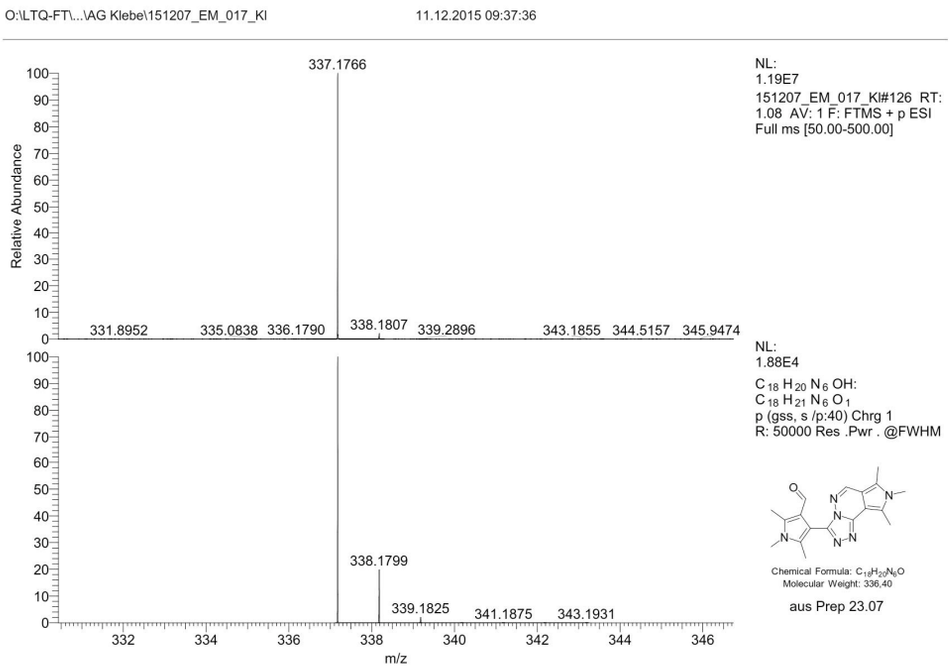


Figure S7.23. ESI-HRMS Spectrum of **3**.

ESI-MS and HR-MS spectra of compound 5

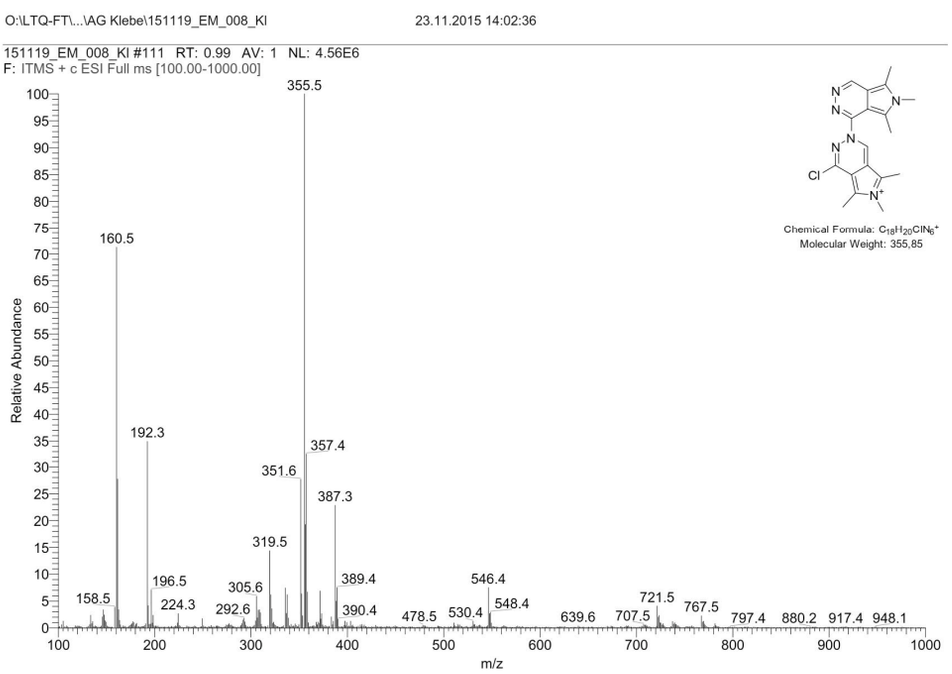


Figure S7.24. ESI-MS Spectrum of **5**.

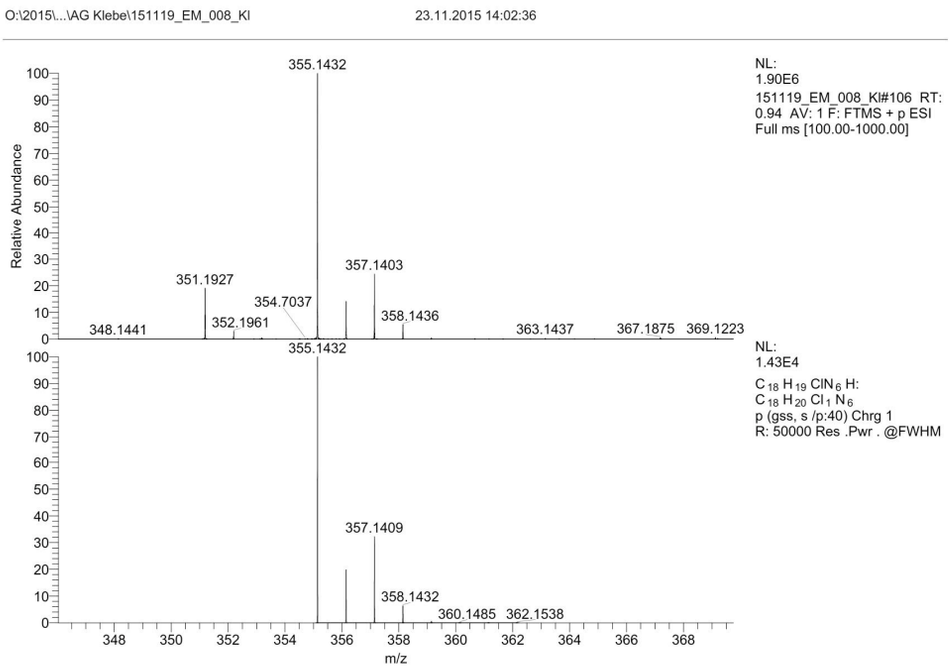


Figure S7.25. ESI-HRMS Spectrum of **5**.

ESI-MS and HR-MS spectra of compound 6

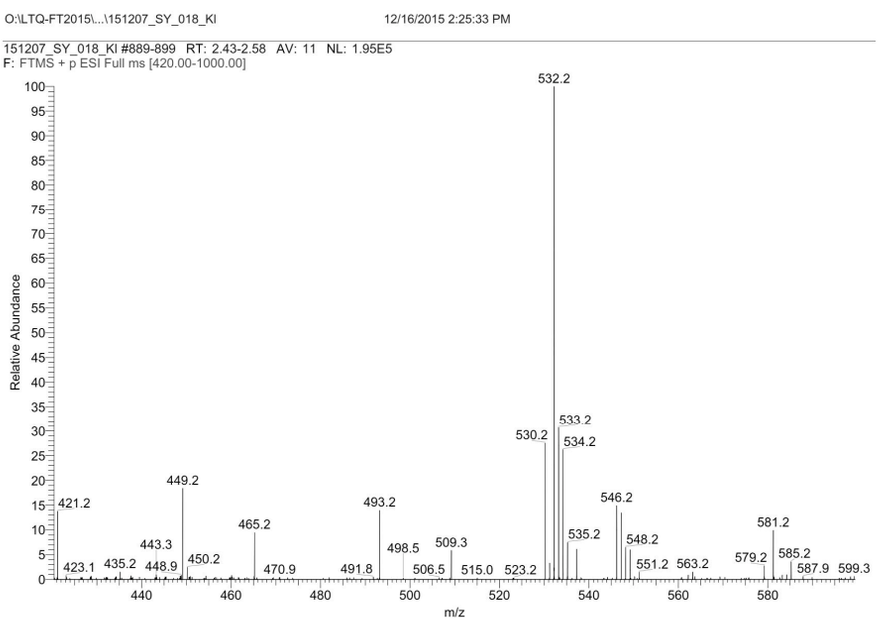


Figure S7.26. ESI-MS Spectrum of **6**.

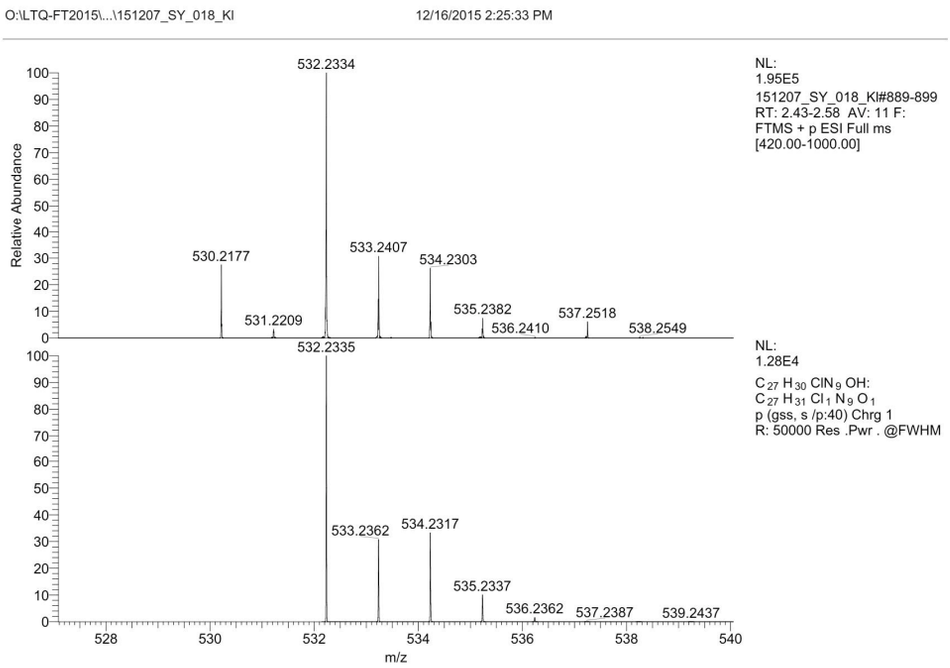


Figure S7.27. ESI-MS Spectrum of **6**.

ESI-MS and HR-MS spectra of compounds 8 and 9

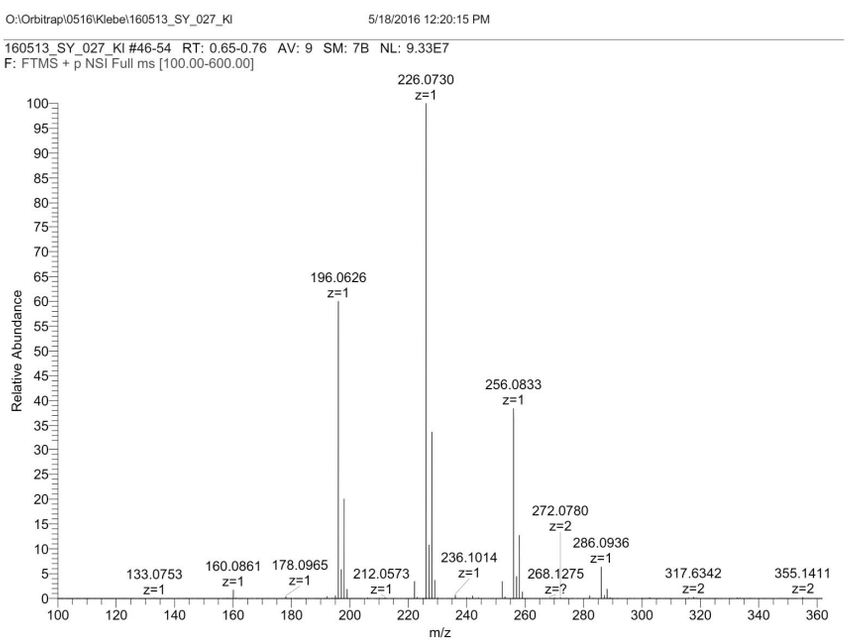


Figure S7.28. ESI-MS Spectrum of **8** and **9**.

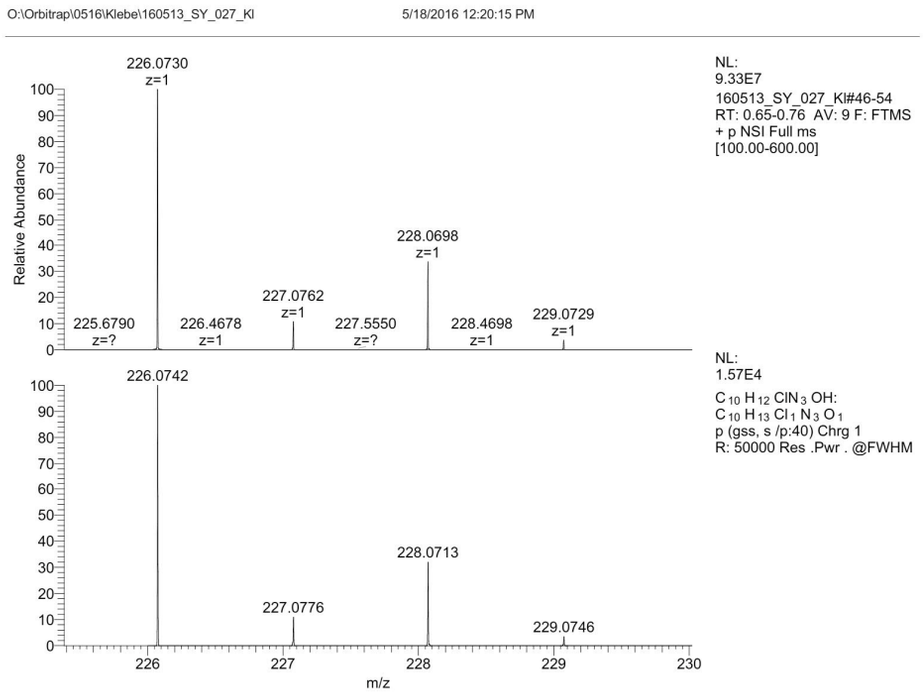


Figure S7.29. ESI-HRMS Spectrum of **8**.

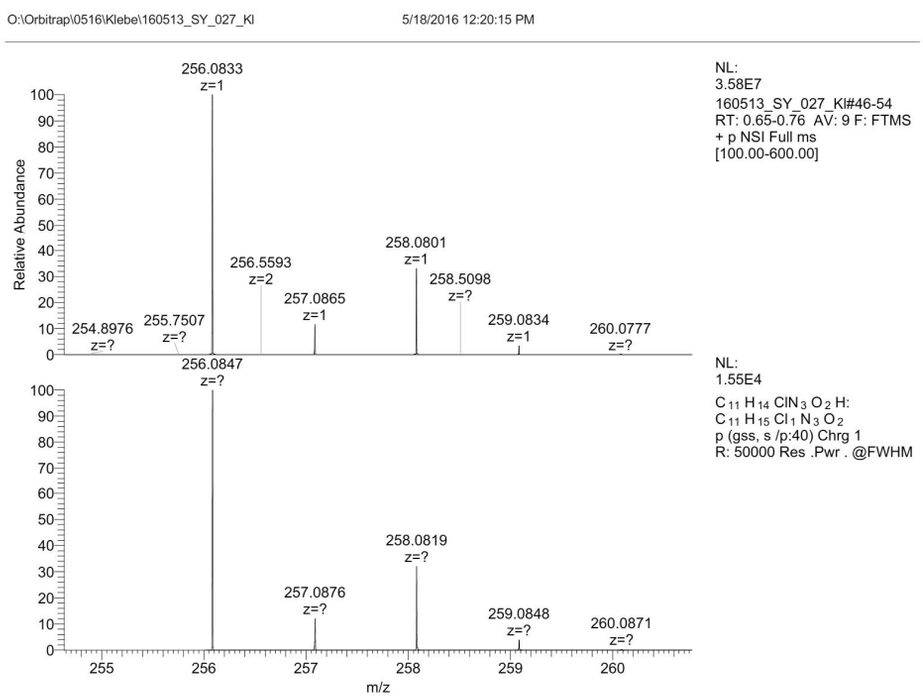


Figure S7.30. ESI-HRMS Spectrum of **9**.

MS/MS Spectra of compound 2

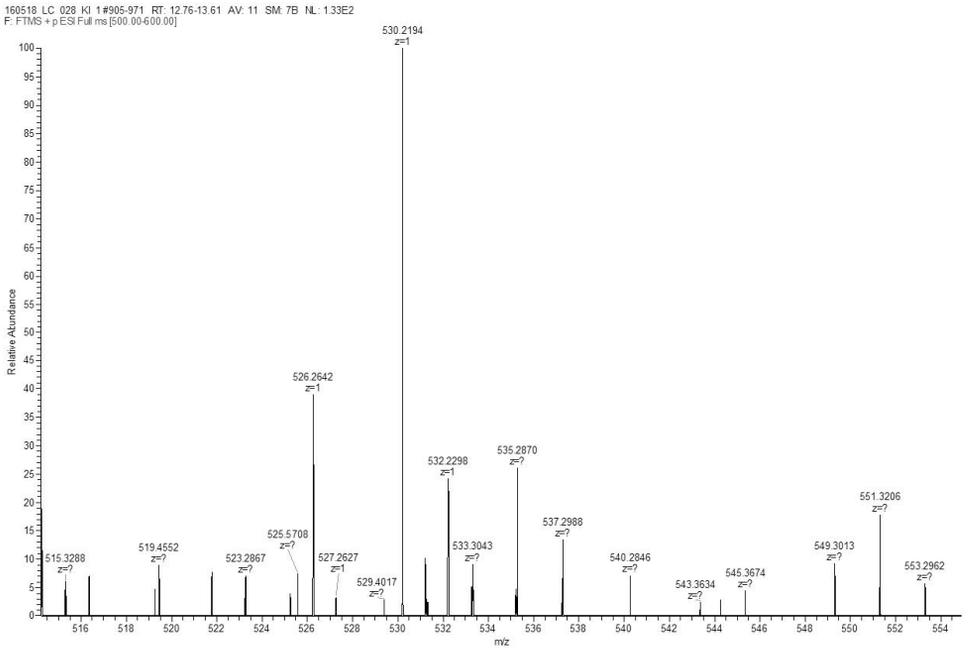


Figure S7.31. MS/MS Spectrum of **2**.

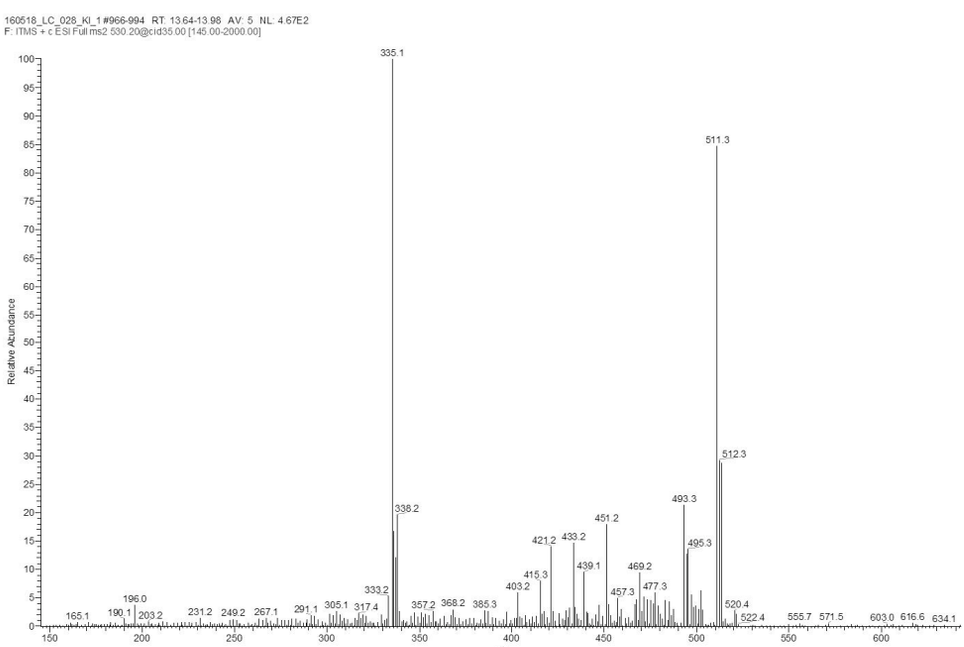


Figure S7.32. MS/MS Spectrum of **2**.

MS/MS Spectra of compound 4

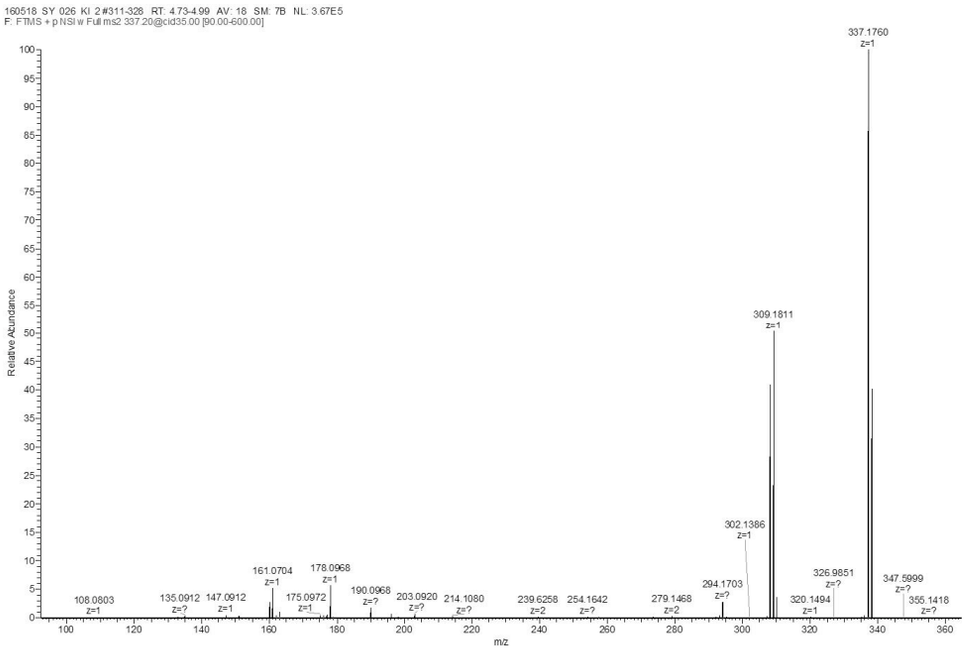


Figure S7.33. MS/MS Spectrum of **4**.

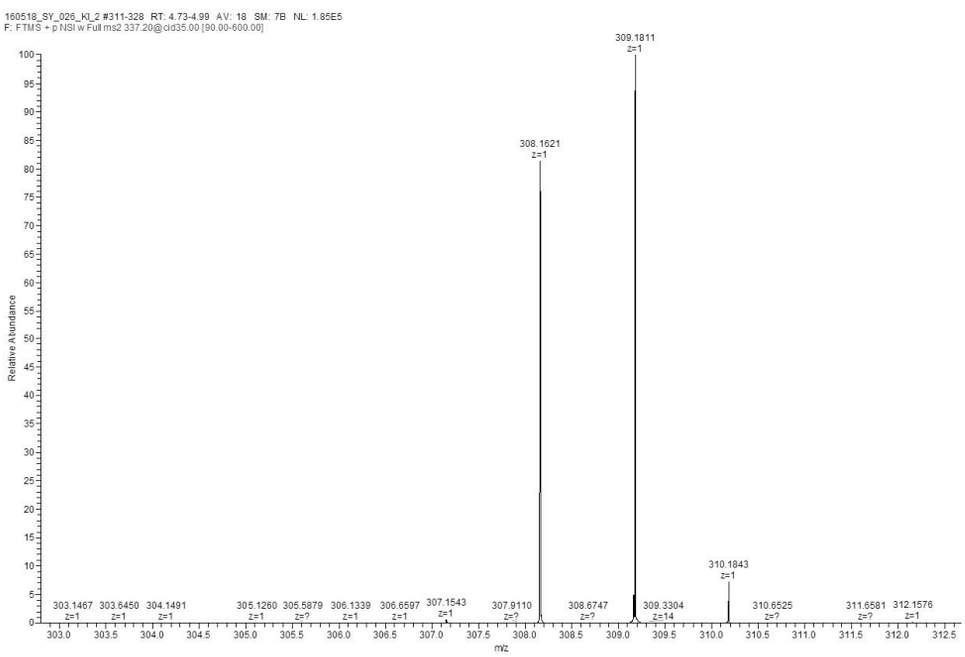


Figure S7.34. MS/MS Spectrum of **4**.

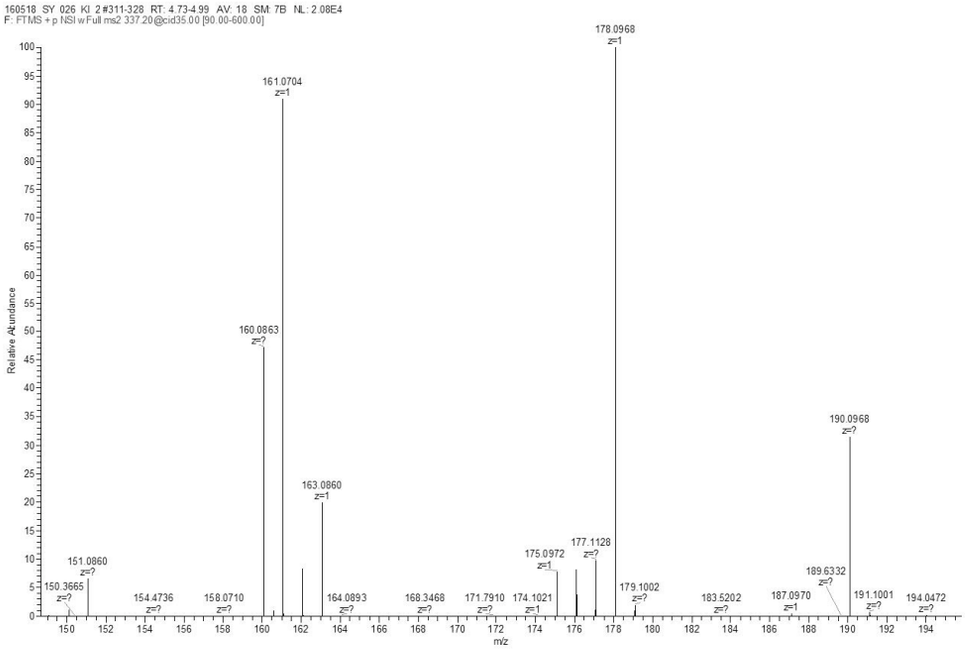


Figure S7.35. MS/MS Spectrum of **4**.

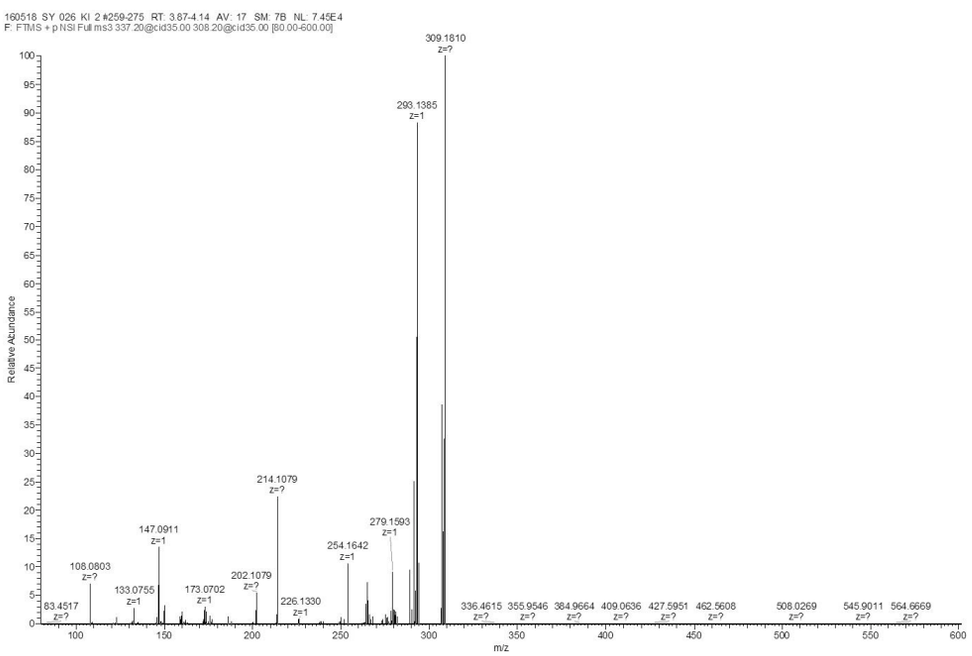


Figure S7.36. MS/MS Spectrum of **4**.

MS/MS Spectrum of compound 5

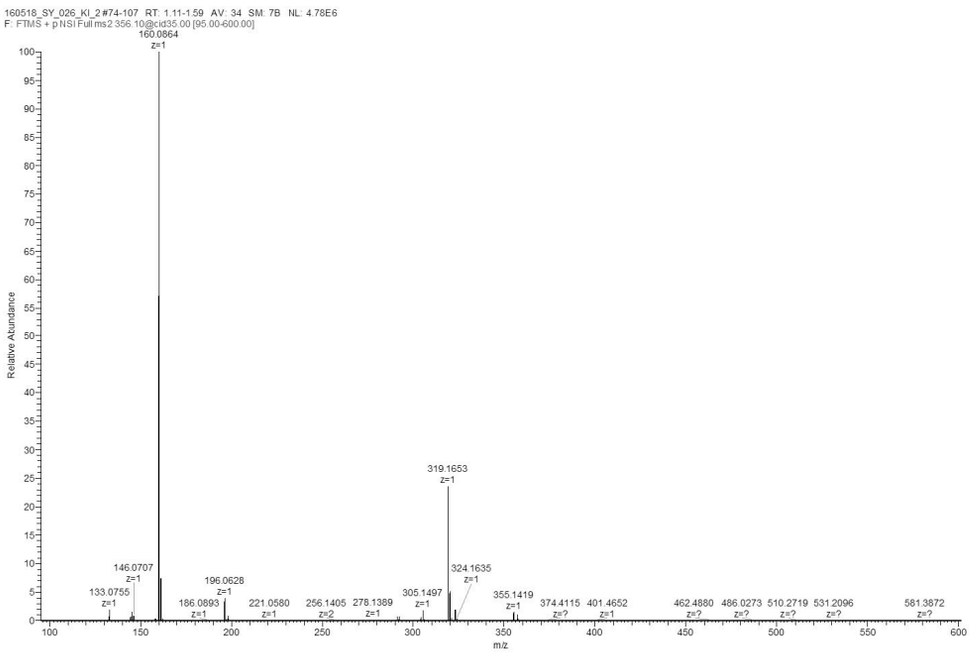


Figure S7.37. MS/MS Spectrum of **5**.

MS/MS Spectra of compound 6

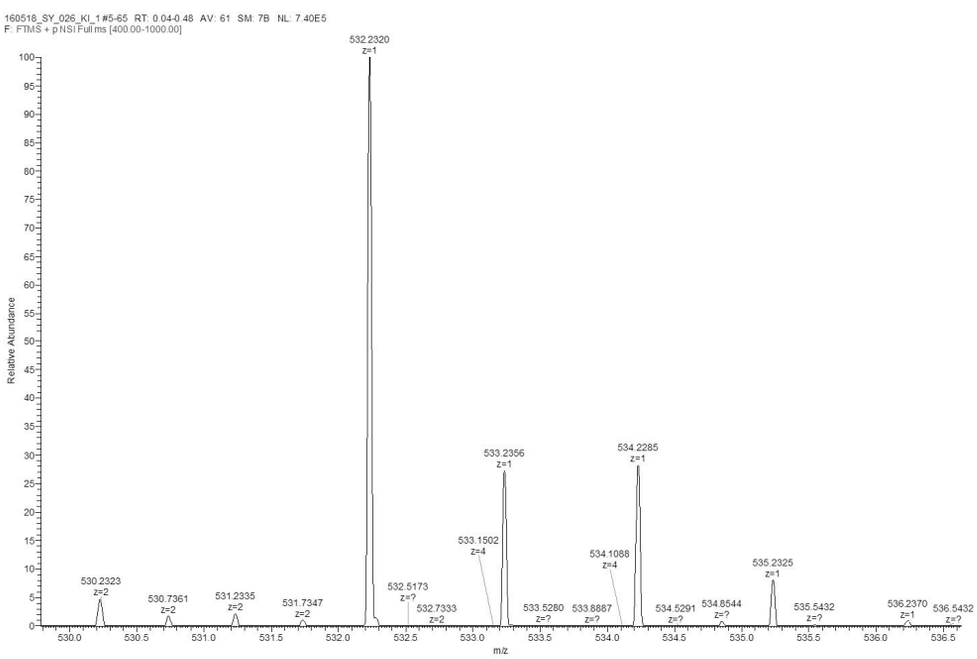


Figure S7.38. MS/MS Spectrum of **6**.

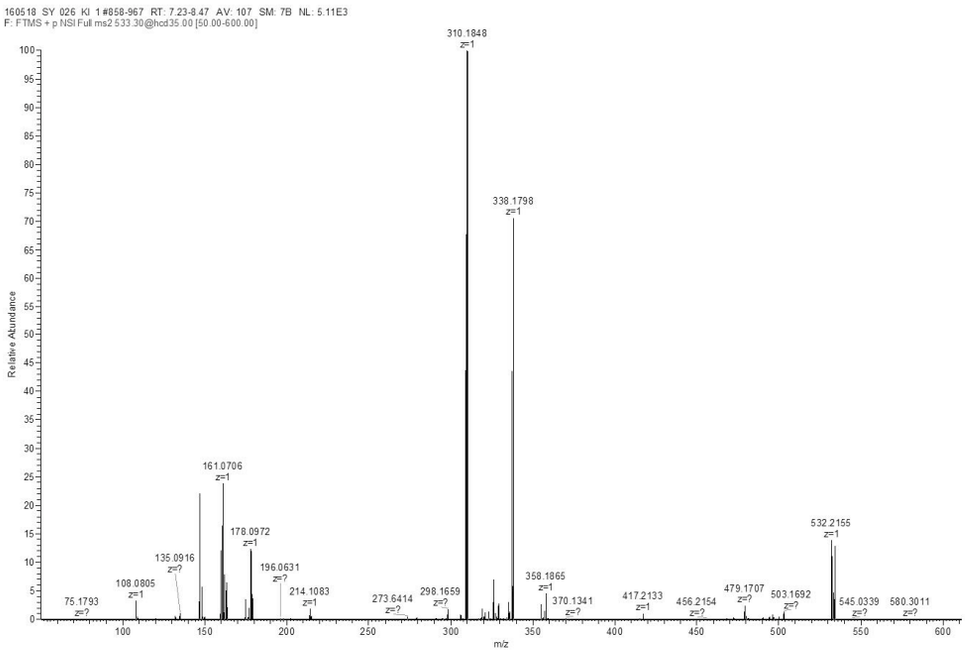


Figure S7.39. MS/MS Spectrum of **6**.

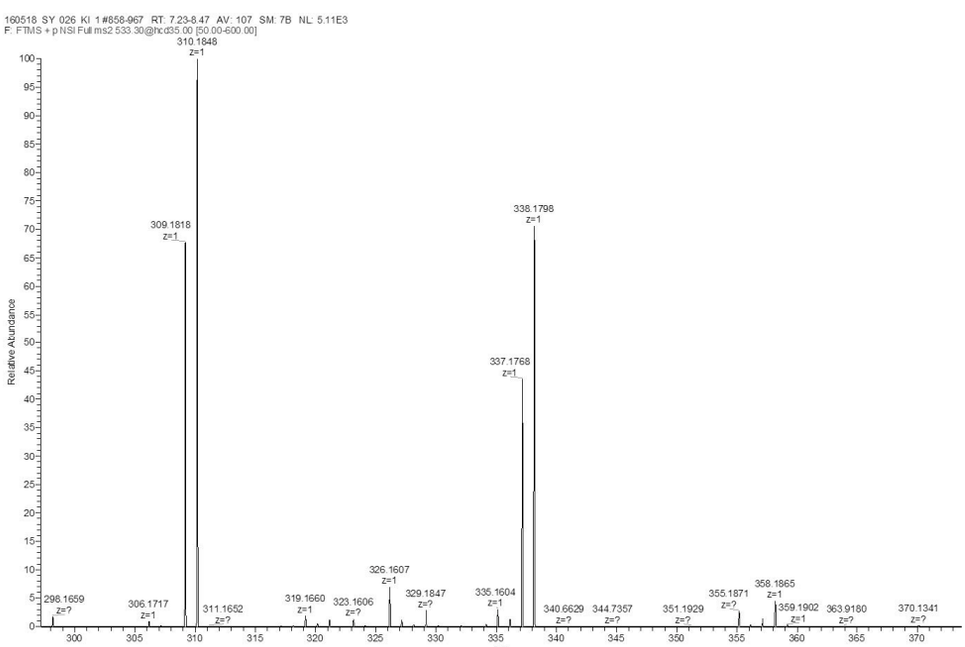


Figure S7.40. MS/MS Spectrum of **6**.

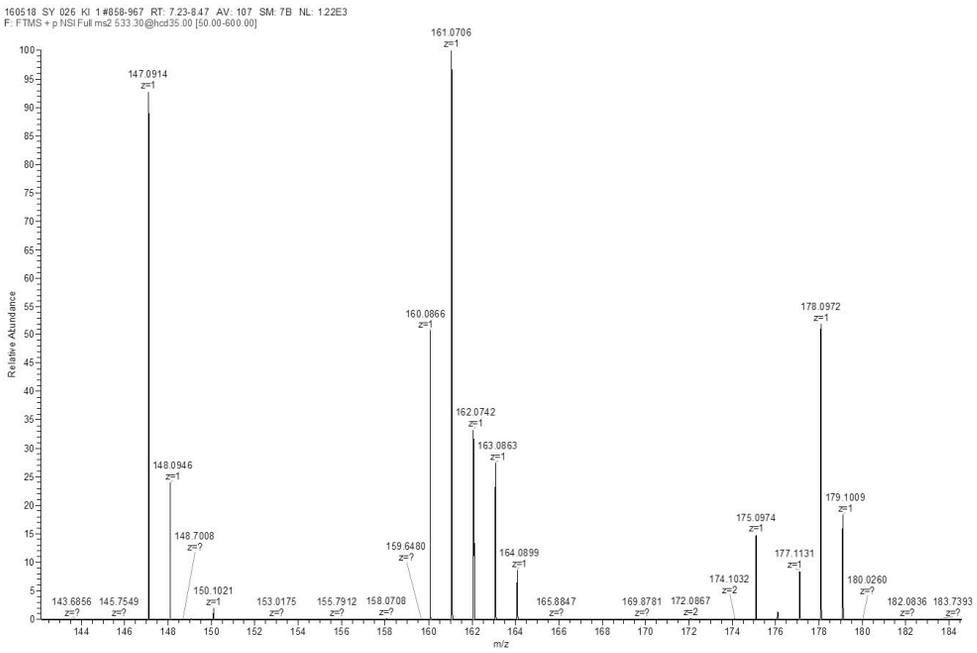


Figure S7.41. MS/MS Spectrum of **6**.

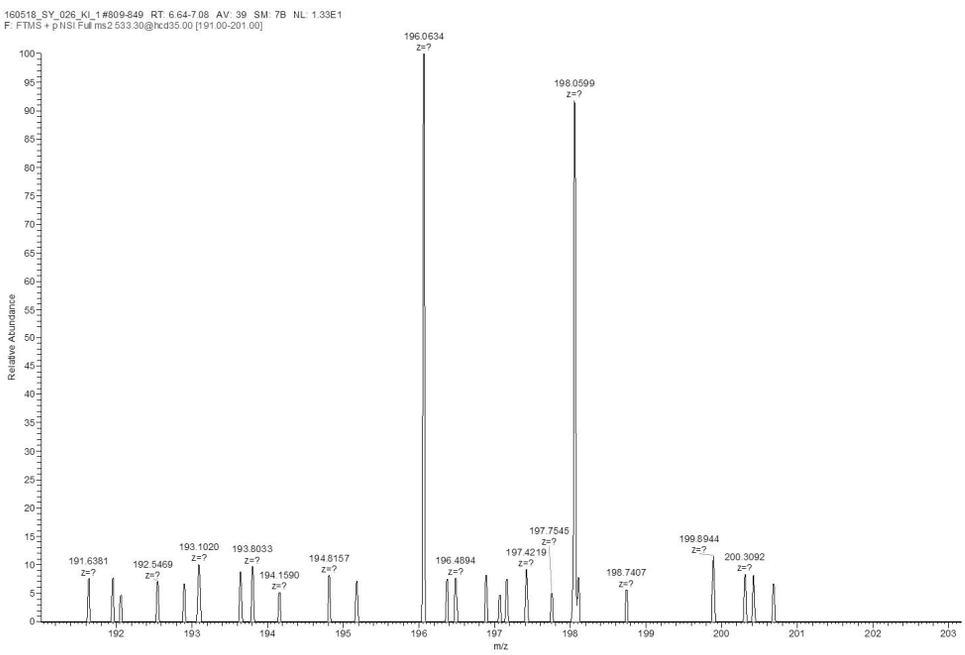


Figure S7.42. MS/MS Spectrum of **6**.

ESI-MS and HR-MS spectra of the crude reaction mixture from the conversion of 5 to 4

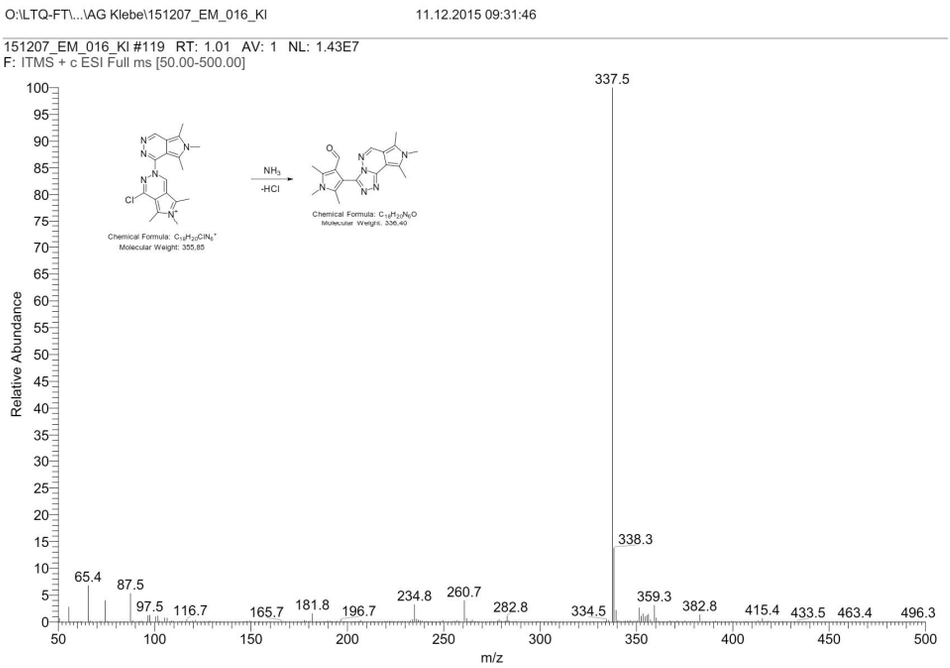


Figure S7.43. ESI-MS Spectrum of the crude reaction mixture from the conversion of **5** to **4**.

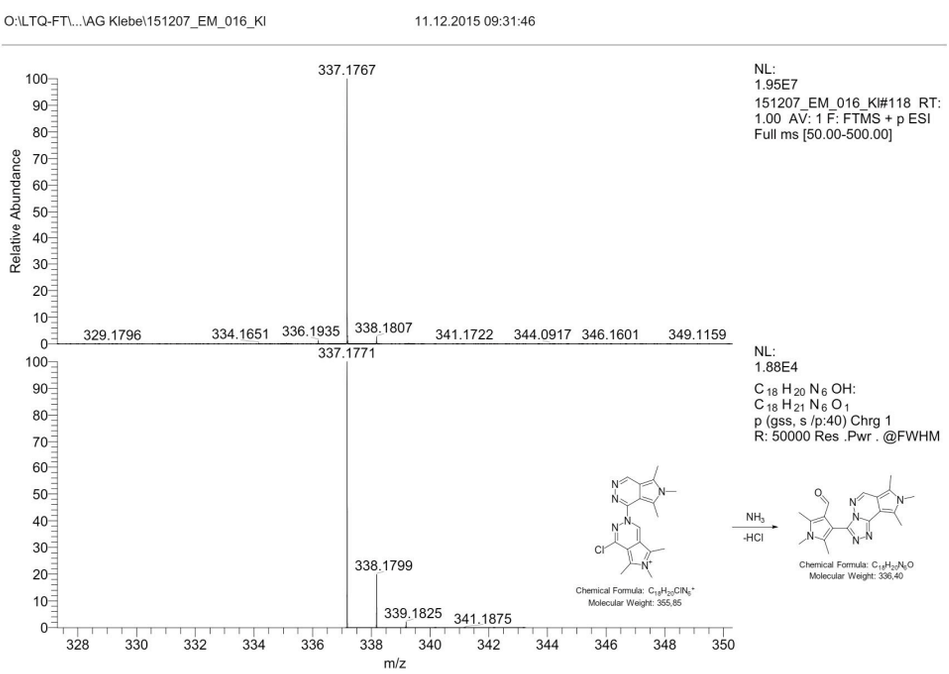


Figure S7.44. ESI-HRMS Spectrum of the crude reaction mixture from the conversion of **5** to **4**.

Chapter 8

Discussion

8.1 Introductory Remark

The previous chapters were taken from self-contained articles in scientific journals. Due to the independent nature of the publications, the compound numbering is ambiguous when the articles are combined to a coherent thesis. In the summarizing discussion in Chapter 8, the compound numbering is expanded to enable a unique identification of each compound. For this, the chapter number is added as a prefix in front of the compound number used in the publication. Thus, the identifier **1.1** refers to compound **1** from the first chapter of this thesis. For a better overview, a summary of relevant compounds is given in each section.

8.2 Synthesis of Phosphonamidate Peptides as Inhibitors of TLN

The application of phosphonamidates in biomedical research has been restricted by their synthetic inaccessibility and their inherent instability under certain experimental conditions. In contrast to the formation of carboxamide peptide bonds, the P–N coupling reaction represents a significant synthetic challenge. The most commonly applied reagents for the activation of the phosphonic acid coupling partner are the phosphonium reagent PyBOP, as well as thionyl chloride and oxalyl chloride (Scheme 1.1). While the use of PyBOP routinely gives good results for the coupling reaction, the method has a significant drawback that is often not considered during the planning phase of a synthesis. During the course of the reaction, equimolar amounts of the byproduct tris(*N*,*N*-tetramethylene)phosphoric acid triamide are formed. After the reaction is complete, this compound has to be separated from the phosphonamidate ester product. Usually this task is easily accomplished by a chromatographic purification step. However, for many of the phosphonamidate peptides that were synthesized in the context of this thesis, a chromatographic purification using silica gel lead to a complete decomposition of the product. This finding was attributed to the acidity of the stationary phase and persisted even with the use of basic modifiers like NEt_3 in the mobile phase. Therefore, thionyl chloride was chosen as the activation reagent for phosphonamidate coupling. The side products of thionyl chloride activation, SO_2 and HCl, can easily be removed by distillation. After a simple aqueous workup, which removes residual starting material and amine bases used for the coupling, the protected phosphonamidate peptide can be obtained in sufficiently pure form.

For this reaction, the stoichiometry, order, and rate, in which the reagents are added to the reaction mixture, was found to be crucial. During the preactivation step of the phosphonic acid, it is important to maintain an excess of the activation reagent over the phosphonic acid in order to suppress the formation of inactive pyrophosphonate intermediates. Therefore, a dilute solution of the phosphonic acid was slowly added to a highly concentrated solution of

two equivalents of $SOCl_2$. With this method, the desired phosphonamidates could routinely be obtained in a high yield.

The most commonly employed protection strategy for phosphonamidate peptides utilizes methyl or ethyl esters for the central phosphonic acid and the *C*-terminal carboxylic acid functions. Saponification using aqueous alkali hydroxide solutions then gives the deprotected inhibitors as lithium or sodium salts. These highly polar compounds are then purified by preparative reversed-phase HPLC. During the course of this process, we routinely struggled with very low recovery rates of the desired inhibitors, which could be attributed to an evident hydrolysis of the P–N bond during the HPLC purification. Even though the pH of the mobile phase was kept neutral, the deprotected phosphonamidate peptides were found to decompose during the chromatographic separation. The hydrolytic instability of the deprotected inhibitors necessitated an adaptation of the synthetic strategy. The hydrolytically labile methyl ester protecting groups were replaced with allyl esters, which can be cleaved via a palladium catalyzed deallylation reaction, using diethylamine as a nucleophilic scavenger. The side products of this reaction can be easily removed by distillation and a subsequent aqueous workup procedure, which gives the deprotected inhibitors as diethylammonium salts. Since the resulting peptides proved to be highly hygroscopic, they were transformed into their respective lithium salts by the addition of LiOH during the aqueous workup step. Alternative purification methods were evaluated in order to avoid the HPLC separation procedure that was previously found to decompose a significant amount of the product. For this task, SPE (solid phase extraction) seemed to be a suitable substitute. While this technique still relies on a chromatographic separation principle, the exposure of the analyte to mobile and stationary

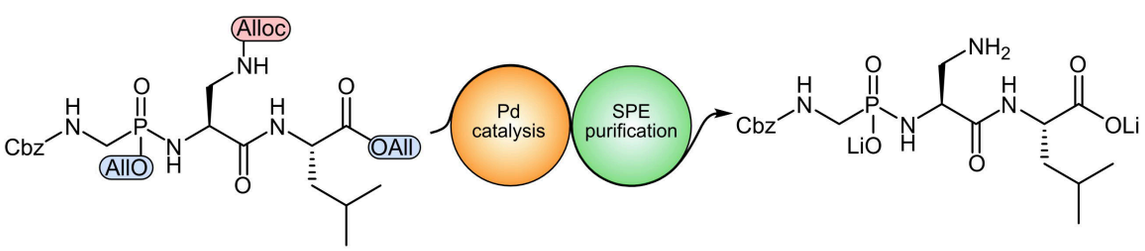


Figure 8.1. Graphical abstract: An allyl protection and improved purification strategy enables the synthesis of functionalized phosphonamidate peptides. Reproduced from [183].

phases is reduced to a minimum. With polymeric DPA-6S resin and cyano or amino modified silica gel phases, no retention of the products or contaminants was observed. The use of zwitterionic ammonium sulfonic acid phases in hydrophilic-lipophilic interaction chromatography (HILIC) proved to be efficient, yet, hydrolysis of the surface modification and the necessity for buffered mobile phases prevent an effective use of this technique. The most success can be achieved with a C_{18} modified silica gel phase. SPE cartridges with this surface modification show little retention of the desired inhibitors, but hydrophobic contaminants are efficiently retained, so that the phosphonopeptides can be obtained in excellent purity.

The developed allyl ester/alloc protection strategy in combination with a SPE purification method represents an operationally simple protocol for the synthesis of phosphonamidate peptides. Using the procedure reported in Chapter 2, 17 inhibitors were successfully synthesized with high purity. Most notably, this procedure allows the incorporation of polar and nucleophilic functional groups, such as alcohols and amines. In addition, an inhibitor containing a hydrolytically labile ester group can be synthesized, which has been proven to be a challenging task with the commonly applied strategy [105]. The excellent purity of the products allows their use in biophysical experiments, such as ITC, that rely heavily on an accurate knowledge of ligand concentrations.

8.3 Understanding Solvation Effects in the Binding of TLN Inhibitors

8.3.1 Rational Design of Water Networks

Upon formation of a protein–ligand complex, the newly generated surface necessitates a rearrangement of the surrounding water molecules in the hydration shell. The shape of this surface has a major impact on the shape of the resulting water structure, which in turn contributes to the binding affinity of the ligand. Solvent effects like this have generally been neglected in the design process of a new drug.

In the study presented in Chapter 3, an attempt is made to include a consideration of this contribution in the design of a ligand for the metalloprotease TLN. Based on a fairly well optimized ligand from a previous study (**3.1**), a series of inhibitors is designed that optimize the resulting water network in the first layer of hydration around their hydrophobic P_2' substituents, which address the shallow and solvent exposed S_2' pocket of TLN (**3.2**, **3.3**). As a control, a set of the respective diastereomers is included in the series (**3.4**, **3.5**). The reversal of the P_2' stereocenter is believed to sufficiently disrupt the water networks, so that its contribution to the thermodynamic binding profile can be evaluated.

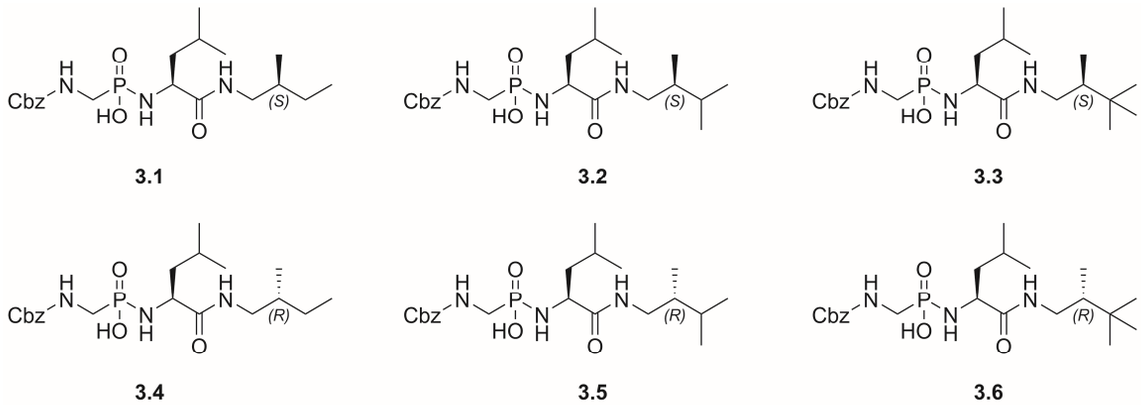


Figure 8.2. Overview of relevant compounds from Chapter 3.

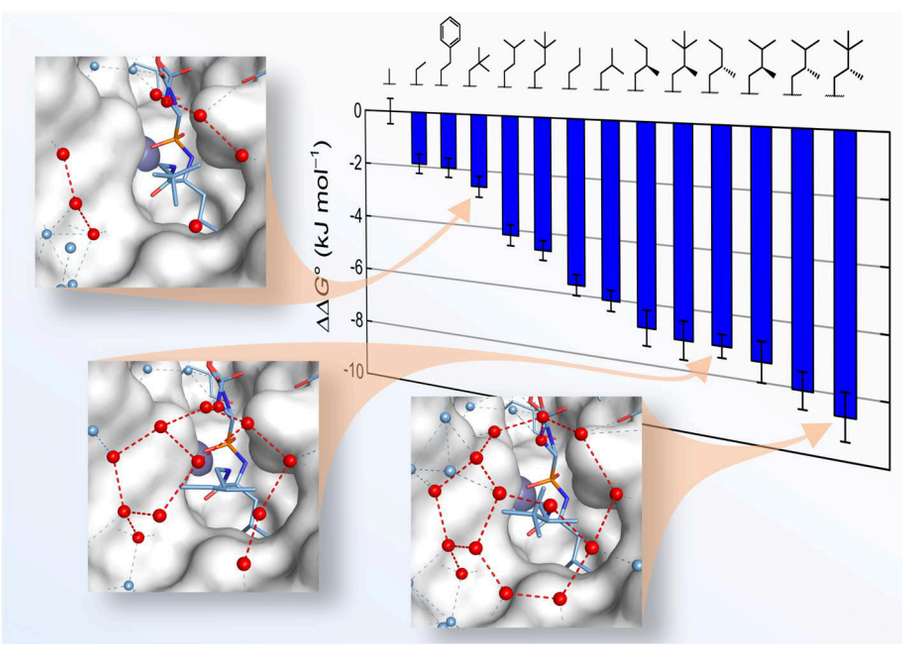
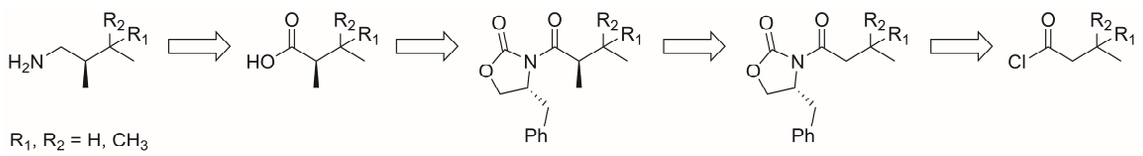


Figure 8.3. Graphical abstract: Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein-Bound Ligands. Reproduced from [16].

An MD simulation, using an explicit solvent approach with a significant damping of the protein movement turns out to be suitable to predict water network formation. This method reveals the most beneficial water organization for ligand **3.3** (Figure 3.11).

The synthetic challenge in the preparation of the envisioned inhibitor series is the introduction of the additional stereocenter of the aliphatic amine in the P_2' position. Not many synthetic methodologies are available for the direct generation of aliphatic β-chiral amines.



Scheme 8.1. Retrosynthetic analysis of chiral alkylamines used in the preparation of TLN inhibitors **3.2**–**3.6**.

While the desired amino building blocks have been synthesized previously by utilizing an enantioselective hydroboration procedure, the reported enantiomeric purities are not sufficient for the intended purpose [291–293]. The copper catalyzed enantioselective hydroamination protocol, reported by Zhu and Buchwald, which generates the desired amines with a much better enantioselectivity [294], was not available at the time of this work. Additionally, neither a crystallographic resolution of the racemic mixture, nor a chromatographic separation of a leucine adduct of the desired molecules proved to be successful. Therefore, the stereogenic center is formally shifted to the α-position of the compounds by a change of the oxidation state of the terminal carbon center. The corresponding α-chiral carboxylic acids are accessible by a diastereoselective enolate alkylation utilizing Evan's chiral oxazolidinone auxiliary. This procedure enables the synthesis of the desired inhibitors starting from the commercially available carboxylic acid derivatives (Scheme 8.1, Scheme 3.1).

From the crystal structures of TLN in complex with the inhibitors it is evident that the MD prediction of the solvation patterns around the P_2' substituents has been very accurate, demonstrating that this approach can be used for the evaluation of hydration propensities of a ligand series without any *a priori* knowledge about the details of binding (Figure 3.11). The crystallographic analysis of the water structure (Figure 3.4) shows that the favorable hydration pattern of the original ligand **3.1** is retained for its *iso*-propyl homolog **3.2** and can even be improved for the *tert*-butyl derivative **3.3**. In contrast to that, the (*R*)-configured epimers **3.4**, **3.5** and **3.6** display a much reduced stabilization of the surrounding water shell. This is especially indicative for **3.4,** which loses the ability to support the favorable five-membered ring water arrangement that is present in the other structures, due to a flip of the orientation of the P_2' chain. Furthermore, an evaluation of the crystallographic *B*-factors reveals a more pronounced stabilization of the water molecules in the first solvation layer for the (*S*,*S*)-configured inhibitors in comparison to their (*S*,*R*)-configured analogues (Figure 3.5).

The thermodynamic binding profiles of the inhibitors were subsequently analyzed by ITC (Figure 3.8) and observed differences in the thermodynamics of binding correlated to changes in the respective crystal structures (Figure 3.4). In comparison to ligand **3.1**, its homolog **3.2**

shows a slightly increased entropic benefit, which can be attributed to a more favorable desolvation entropy, owing to the larger hydrophobic surface that is buried upon binding. Additionally, the observation of two distinct conformations in the crystal structure implies an intrinsic disorder of the aliphatic P_2' chain, which might give an entropic benefit. This effect is superimposed by an enthalpic loss that might originate from a disorder in the water structure, weakening solvent–solvent interactions.

In the structure of **3.3**, additional water molecules are recruited to the complex (Figure 3.4C). An extensive network of multiple fused water polygons is observed. This inhibitor also shows the highest affinity of the series. However, the free energy gain determined for this compound originates from a beneficial entropic term compared to its *iso*-propyl analog **3.2**. This observation was credited to an increase of the buried hydrophobic surface area for the *tert*-butyl group. While the enlargement of the hydrophobic portion results in a destabilization of the water network with a concomitant loss in enthalpy in the case of **3.2**, the shape of the hydrophobic surface generated by the *tert*-butyl in **3.3** group allows an optimal stabilization of the solvation layer. This compensates for the desolvation cost of the larger hydrophobic residue, resulting in a more beneficial free energy of binding.

The crystal structures of the (*R*)-configured ligands (Figure 3.4E/F) reveal a significant increase of the *B*-factors of the water molecules in the network surrounding the P_2' moiety. The inversion of the stereogenic center results in an increased steric demand of the (*R*)-configured side chains that is sufficient to disrupt the solvation pattern of the hydration shell. As a result, the expected entropic gain upon burial of an increased patch of hydrophobic surface is entirely compensated by an enthalpic loss, nullifying the free energy balance. Thus, the gain in affinity, which was observed in the (*S*)-configured series, represents an enthalpic advantage from a beneficial stabilization of favorable water structures in the first solvation layer.

8.3.2 The Price of Solvation for Ligands Binding to Protein–Solvent Interfaces

Charged and polar functional groups are often included into a ligand scaffold to improve its solubility or pharmacokinetic properties. When the target protein does not offer a suitable interaction for the newly introduced function, it is commonly placed into a position, from which the polar substituent sticks out into the solvent and remains partially hydrated. The study in Chapter 4 is aimed to investigate the phenomenon of partial desolvation of a charged and polar functional group and evaluate the binding thermodynamics in comparison to an inhibitor featuring a purely hydrophobic group in this position. The crystal structures of TLN in complex with ligands **4.1**–**4.8**, which feature a protonated ammonium group at the terminus of an aliphatic chain of variable length in the P_2' position, reveal that the terminal functional group tends to interact with Asn111 (Figure 4.2). Whenever the length of the aliphatic chain

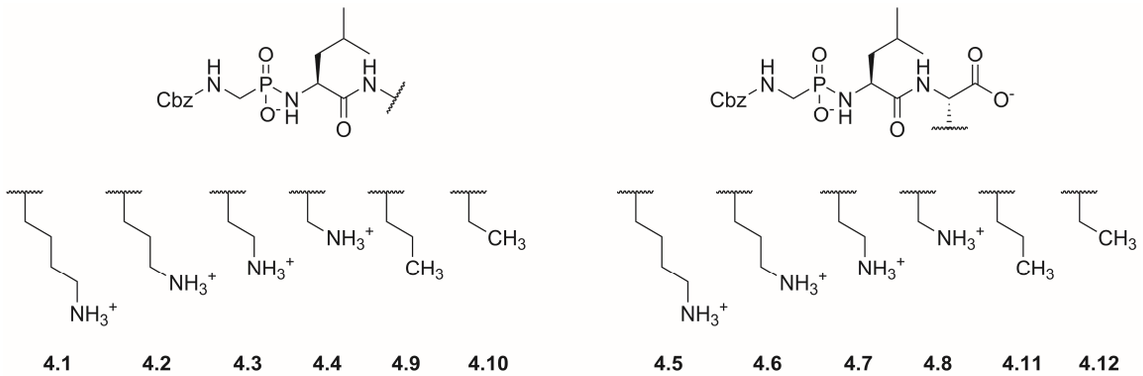


Figure 8.4. Overview of relevant compounds from Chapter 4.

permits, the backbone carbonyl group of this residue accepts a hydrogen bond from the charged moiety. For the ligands with a shorter aminoethyl or aminomethyl group, the interaction distance becomes too large and the hydrogen bond to Asn111 is taken up by a water molecule instead. In these cases, the ammonium groups interact with solvent molecules that are recruited into the complex. The thermodynamic binding profiles are characterized by a steady decrease of affinity upon reduction of the linker length (Figure 4.1). The free energy of binding is split up into enthalpic and entropic contributions that, apart from minor compensations, do not dramatically deviate between the ligands. Only for the shortest ligand (**4.4**), a striking loss in the enthalpic term is recorded. This contribution might originate from an increased cost for the partial desolvation of the charged group, as it is pulled further into the hydrophobic environment of the S_2' pocket. The unfavorable enthalpic term, however, is partially compensated by an entropic benefit. The necessity to assign multiple conformations to the ligand side chain in the refinement model indicates a disorder in the system that might explain the recorded entropic gain. The ligands, comprising a lysine homolog in the P_2' position (**4.5**–**4.8**), adopt similar binding modes when compared to their isostructural analogs. Depending on the linker length, an interaction to either Asn111, or water molecules in the hydration shell is observed. Only for the longest ligand (**4.5**), the terminal ammonium group sticks out into the solvent and is disordered in the crystal structure. The thermodynamic binding profiles within this series also reveal a slight decrease in affinity when the linker length is gradually reduced. This is driven by an enthalpic loss that is partially compensated by an entropically beneficial contribution. For the shortest ligand of this series (**4.8**), a pronounced gain in enthalpy, which is compensated by an entropic loss, can be observed. The enthalpic benefit of this inhibitor might originate from the stabilization of six additional water molecules in the solvation shell around the terminal ammonium group. Consequently, this loss of mobility is also recorded as a penalty in the entropic term. Even though the ligand side chain

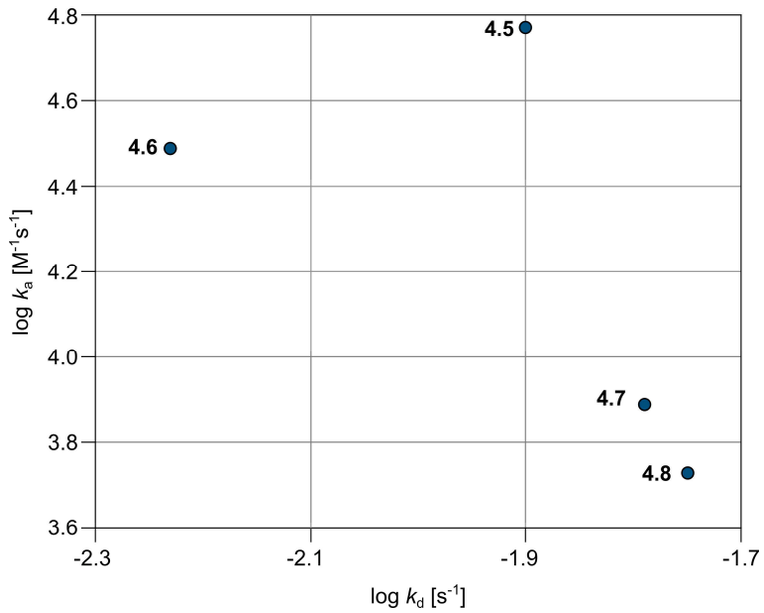


Figure 8.5. Binding kinetic data of **4.5**–**4.8** from SPR measurements.

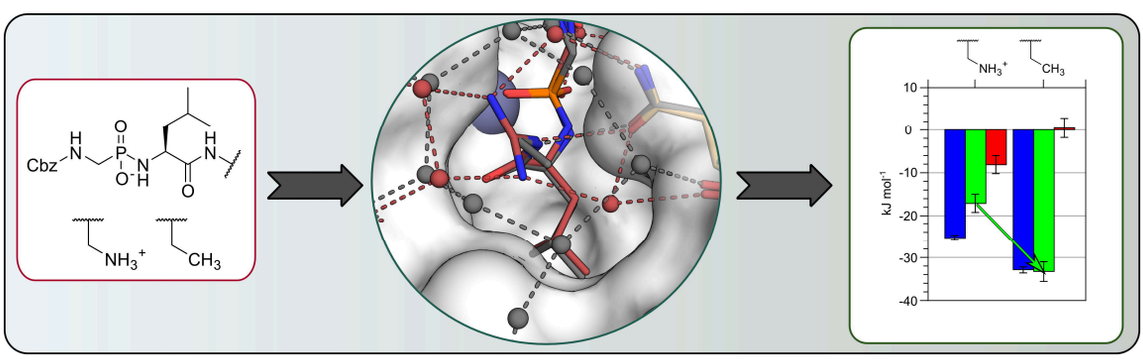


Figure 8.6. Graphical abstract: Paying the Price of Solvation in Solvent-Exposed Protein Pockets: Impact of Solubilizing Polar Groups on Affinity and Binding Thermodynamics in a Series of Thermolysin Inhibitors.

is disordered in the crystal structure, the loss of degrees of freedom in the hydration shell dominates the thermodynamics of binding.

Binding kinetic data recorded by SPR measurements of four compounds (Figure 8.5) reveal a steady decrease in the rate constant of association, while the dissociation rate is mostly constant within the series. Only for the ornithine derivative (**4.6**), a slightly slower off-rate is observed. The association mechanism of inhibitors of TLN is likely governed by a conformational adaption of the protein. During this process, Glu143 is protonated and recedes deeper into the

S_1 pocket. In addition, the catalytic water molecule that is bound to the central zinc ion has to be displaced into the bulk. Since the inhibitor scaffold of the investigated ligands is kept constant, it is conceivable that a significant variation in this part of the process can be ruled out. Most likely, the decrease in k_a with a reduction of the chain length can be attributed to an increasing kinetic penalty for the desolvation of the terminal ammonium functionality.

A comparison of the characteristics of binding with structural analogs that feature a purely hydrophobic substituent in the P_2' position (**4.9**–**4.12**) yields valuable insight. In either case, the apolar compounds are significantly more potent than their hydrophilic analogs. This gain in free energy by a factor of 20–180 in K_d is the result of a dramatic enthalpic penalty that accompanies the inclusion of a charged moiety in the inhibitor scaffold. Probably as a result of the partial desolvation of the polar ammonium group, the respective inhibitors lose potency. An inspection of the crystal structures reveals that the organization of residual solvent molecules around the P_2' groups of the ligands does not favor the polar compounds (Figure 4.3). In all cases, an equivalent or even superior network of water molecules is established if a strictly aliphatic residue is included in this position. This highlights the fact that contributions of the solvent as a third interaction partner in protein–ligand binding have to be considered in the design of a prospective drug molecule. In the case of the inhibitor series investigated here, the partial desolvation of a polar group dominates the interaction, even when it is bound to a mostly solvent-exposed binding pocket.

8.3.3 Exploiting Poorly Solvated Binding Pockets in Drug Design

It has been postulated in literature that the accommodation of water molecules in hydrophobic cavities depends on the shape of the surface, and that insufficiently solvated protein pockets exist in nature [11,14,295]. In the study presented in Chapter 5, the S_1' pocket of TLN is identified as such a poorly hydrated cavity. The ligand that was used in the initial experiments features a glycine residue in its P_1' position (**5.1**). This configuration covers the S_1' pocket, while the lack of a bulky substituent hypothetically allows the accommodation of several water molecules in the residual cavity of > 140 Å^3. In order to obtain a reliable estimation of the content of the enclosed crevice, it is necessary to utilize experimentally determined phase information for the calculation of an electron density map. In comparison to the most commonly applied phasing technique, molecular replacement, experimental phasing does not only avoid bias, but also allows the computation of absolute electron content instead of the usually applied sigma scaled electron density calculation. For this study, a MAD experiment using the anomalous scattering properties of the catalytic zinc ion in TLN was utilized. The resulting electron density map revealed that no water molecules are bound in the S_1' cavity below the inhibitor (Figure 5.2).

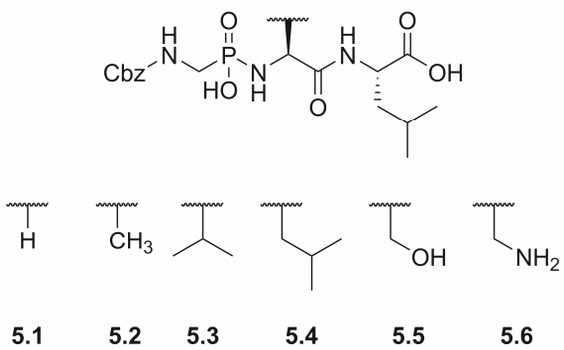


Figure 8.7. Overview of relevant compounds from Chapter 5.

In comparison to three reference pockets in various states of hydration, no electron density that might correspond to a disordered water molecule can be detected in the crystal structure. In a second experiment, crystals of TLN were treated with gaseous krypton and xenon. The noble gasses are found to bind to the hydrophobic S_1' cavity in the presence of the inhibitor (Figure 5.3). From the electron configuration and the crystallographic temperature factors of the xenon and krypton atoms, it can be estimated that, considering the size of the pocket, disordered water molecules with an occupancy as low as 14% should be observable with crystallographic techniques. The results of these experiments suggest that the S_1' pocket of TLN is indeed unsolvated, and thus empty, when no ligand is bound to the protease. This observation raises the question how the accommodation of a ligand portion in the empty space translates into the thermodynamics of binding. To analyze this, a series of ligands was synthesized that feature a hydrophobic substituent, gradually increasing in size, in the P_1' position (Figure 8.7). The introduction of a methyl group alone (**5.2**) results in an affinity gain by a factor of 100 in K_d (Figure 5.1). A further growth of the hydrophobic substituent up to a leucine residue (**5.4**) then yields a further 40-fold increase of potency. The decrease in free energy of binding is mainly driven by a favorable enthalpic contribution, which originates from the reduction of the vacuum inside the S_1' cavity (Figure 5.4).

The results of this study highlight the potential that poorly hydrated protein pockets provide for drug design. When such a cavity can be identified and properly addressed by a ligand, a significant gain in potency can be expected.

8.4 The Origin of Long Residence Time Binding in TLN

The pursuit of long residence time binders has become a prominent objective in drug discovery projects. Molecules with this feature are believed to show a better correlation of preclinical data

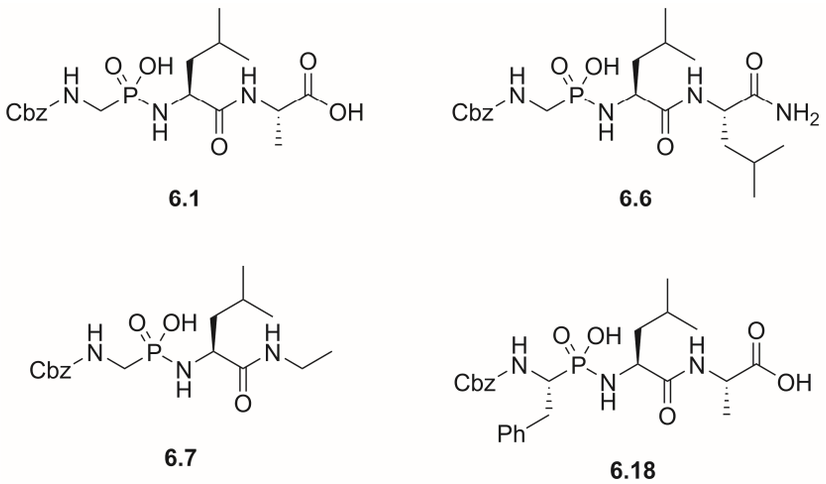


Figure 8.8. Overview of relevant compounds from Chapter 6.

to *in vivo* efficacy and display an increased selectivity against their target in the complex environment of a cell [1,21,23,115]. However, medicinal chemists currently lack the tools for a targeted alteration of the kinetic binding parameters of a ligand, because the underlying mechanisms are not well understood. The metalloprotease TLN has proven to be a valuable model system for the study of protein–ligand interactions. In order to improve our understanding of the molecular determinants of long residence time binding, binding kinetic data of a series of TLN inhibitors was determined by SPR and correlated to crystallographic data of the corresponding protein–inhibitor complex. The investigated compound series contained phosphonopeptides with differing functional groups in their respective P_2' portion. In this position, a charged and polar carboxylic acid, a polar but uncharged carboxamide or an apolar aliphatic residue was accommodated. Representative members of the series are given in Figure 8.8. SPR results revealed that compounds containing a terminal carboxylic acid displayed the highest dissociation rates throughout the series. This value was reduced for the carboxamide compound, which showed an intermediate residence time. For the aliphatic compounds, the fastest dissociation kinetics within the investigated compound series was found (Figure 6.1).

Since the binding mode of the inhibitors is conserved for all members of the series (Figure 6.2), the differences in the dissociation rates could be linked to the interaction of the compounds with the side chain of Asn112. This residue is accommodated above the S_1' pocket of TLN and usually assists in the positioning of a substrate during peptide bond cleavage. Inhibitors featuring a carboxylic acid in the P_2' position are able to engage in a strong, charge assisted hydrogen bond to this residue. This results in a residence time around several minutes. For the carboxamide compound, this interaction pattern is conserved, although the negative charge

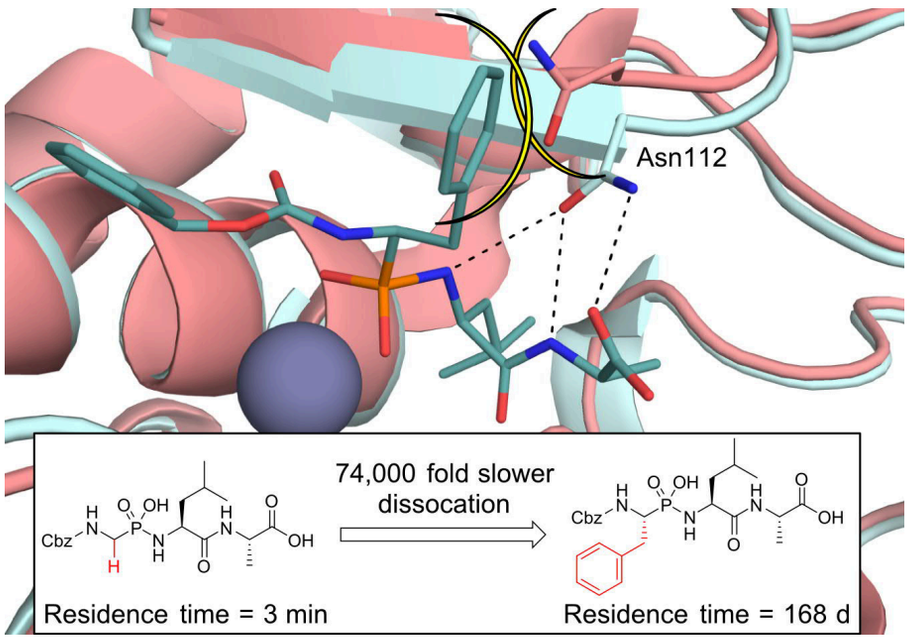


Figure 8.9. Graphical abstract: Elucidating the Origin of Long Residence Time Binding for the Inhibitors of the Metalloprotease Thermolysin. Reproduced from [179].

assisting the interaction is lost, therefore weakening the resulting hydrogen bond. The loss of a polar interaction partner for Asn112 finally results in a significant decrease in residence time to the order of several seconds for the aliphatic compounds (Figure 6.3).

The strong correlation of the residence time to the ability of an inhibitor to interact with the Asn112 side chain gave rise to the hypothesis that this residue might be involved in the rate limiting step of dissociation. During the turnover of a substrate, TLN and related M4 metalloproteases are believed to undergo a hinge-bending motion around their central α-helix [160,255,256]. From the crystal structure of the *apo* form of TLN, it is evident that this conformational adaption includes a rotation of the Asn112 side chain toward the active site of the protein. Supposedly, this rotation is imperative for the release of a bound inhibitor or substrate, since the side chain carboxamide of Asn112 is located directly above the S_1' specificity pocket that usually accommodates a large, apolar residue. This hydrophobic substituent is thoroughly fixed in its position until Asn112 gives way and allows the cleavage product to dissociate from the well-defined and deeply buried binding pocket (Figure 6.4). The experimental evidence suggests that this retro induced fit mechanism is the rate limiting step for the release of a bound substrate or inhibitor, and thus, the dissociation kinetics of TLN.

The paramount function of Asn112 is further underlined by the finding that this residue is highly conserved within the M4 protease family (Figure 6.5). The related family of M13 proteases, which contains several key enzymes involved in the pathophysiology of heart failure,

Alzheimer's disease and certain types of cancer, shows a close structural analogy [259–261]. Furthermore, a mutagenesis study revealed that TLN mutants, in which Asn112 was exchanged against other amino acids, lost their catalytic activity. Only N112D and N112E mutants partly retained their proteolytic capabilities, supposedly due to the close chemical relation of the newly introduced residues to asparagine [262,263].

In 1983, Bartlett and Marlowe reported binding kinetic data of the TLN inhibitor ZF^{P}LA (**6.18**) [264]. This compound is chemically related to the phosphonopeptides studied for this work, but features a (*R*)-benzyl substituent in the P_1 position. ZF^{P}LA is the most potent known inhibitor of TLN (K_d = 68 pM) with a remarkable residence time of 168 days. From the crystal structure of this molecule in complex with TLN it is apparent that the binding mode of the Phe^{P} compound ZF^{P}LA shows a significant difference to that of the less potent Gly^{P} inhibitors (Figure 6.7). The P_1 benzyl moiety is oriented toward the solvent above the S_1 pocket and placed within van der Waals distance of Phe114. Most importantly, this arrangement blocks the outward movement of Asn112, as evident from a superposition of the open-state structure. Since the conformational adaption of Asn112 was found to be a prerequisite for the dissociation of a bound inhibitor, this steric restriction results in a functionally irreversible inhibition of the enzyme. This example demonstrates the effect that an inhibition of the retro induced fit mechanism in TLN can represent. The conservation of this structural motif throughout the M4 and M13 protease family implies that this mechanism might be exploited for the development of inhibitors against other pharmaceutically relevant targets.

8.5 A False Positive in a Fragment Screening Campaign

From an elaborate fragment screening campaign of a library of 361 entries, **7.1** emerged as one of the most potent endothiapepsin binders in multiple orthogonal assays. It is quite surprising that the molecule found to bind to the active site of EP by X-ray crystallography, does not resemble this compound, but instead indicates that a much larger species is accommodated within the binding cleft. The purity of the substance used for the crystallographic experiments was verified by NMR and MS, so it seems apparent that the unknown molecule is generated under the conditions of the experiment.

Monitoring a sample of the fragment in solution by HPLC reveals that a number of different species are formed simply by exposing the compound to a protic solvent (Figure S7.1A). The intermediates from this reaction were isolated by preparative HPLC and their chemical composition was analyzed by NMR and MS. Based on the structure of the intermediates, a reaction mechanism for the observed degradation of the sample is proposed (Scheme 7.1). A substitution of the halogen substituent in the chloropyridazine moiety of **7.1** by an imine nitrogen of a second molecule, initiates the reaction cascade. The charged dimer **7.5** is the

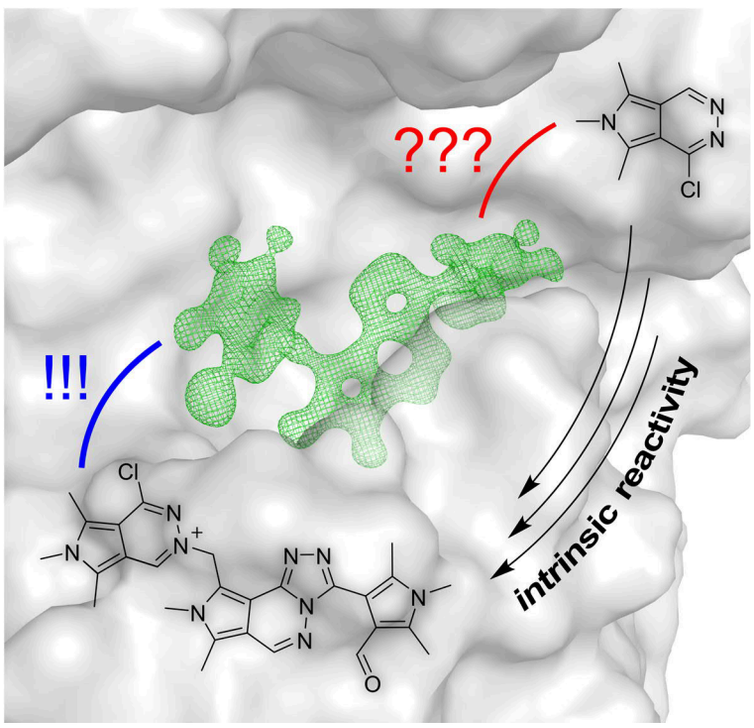


Figure 8.10. Graphical abstract: Watch Out for the Red Herring: Surprising Reactivity of a Fragment Results in Biological Activity. Reproduced from [302].

product of that reaction. In an alternate pathway, a solvolysis of **7.1** to its methoxy derivative **7.3** is observed when methanol is used as a solvent. In the presence of hydroxide ions, the central iminium function in the charged dimeric intermediate **7.5** is hydrolyzed to give aldehyde **7.7** as a hypothetical intermediate. Under formation of a central 1,2,4-triazole heterocycle, this species reacts in an intramolecular substitution reaction to **7.4**. A trimeric compound, which originates from an addition of **7.1** to the aldehyde function of **7.4**, is also observed. While none of the isolated compounds corresponds to the unknown binder observed in the crystal structure, it seems likely that it is formed from **7.4**, whose heterocyclic core structure shows a close resemblance to the connectivity inferred from the original electron density. Based on this finding, structure **7.2** could be assigned to the compound found in the binding site of EP. In addition, a molecule with the corresponding molecular mass was found in minute concentration in the raw reaction mixture by LC-MS.

Fortunately, one of the crystal structures of EP in complex with **7.2** could be refined to a resolution of 1.05 Å. This allows an unrestrained least-squares refinement of the geometric parameters of the ligand. The bond lengths, torsion, and dihedral angles are in line with the proposed structure of the molecule (Figure S7.4, Figure S7.5). The observed electron density

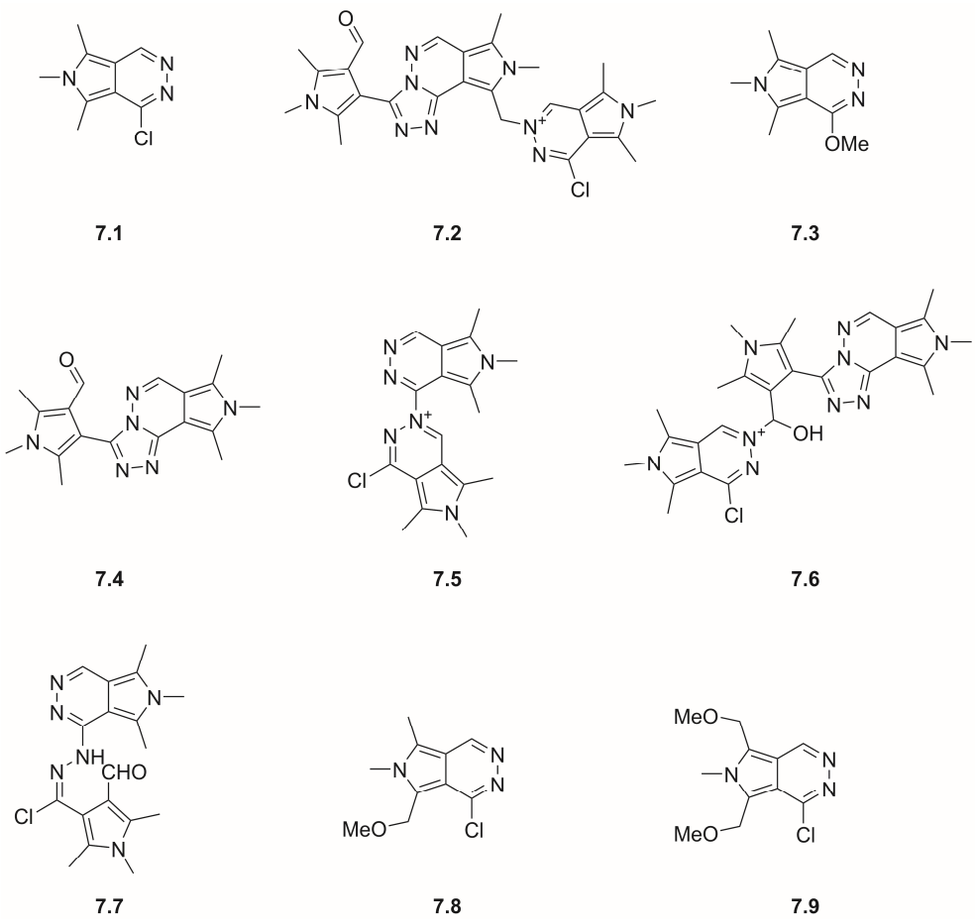


Figure 8.11. Overview of relevant compounds from Chapter 7.

cannot be explained by the presence of two separate molecules, which precluded the simultaneous or subsequent binding of **7.1** and **7.4** with reduced occupancy.

For a hypothetical synthesis of the target compound from **7.1** and the **7.4**, a C-H activation of one of the methyl groups would be a mechanistic requirement. Since the methyl centers in related 2-methyl pyrroles and 2-methyl indoles are known to react under catalytic or anodic oxidation, as well as under photochemical activation [278,280,281,296], it is reasonable to assume that a single electron transfer reaction is involved in the formation of **7.2**. An exposure of the original fragment to the irradiation of a medium pressure mercury lamp in a photochemical reactor gives a nearly quantitative transformation to the methoxylated derivatives **7.8** and **7.9** (Scheme 7.1). This finding might indicate that the methyl positions in chemically related compounds, such as **7.4**, are susceptible to a photochemical activation. The reaction product **7.2**, however, can also be detected by LC-MS when a sample

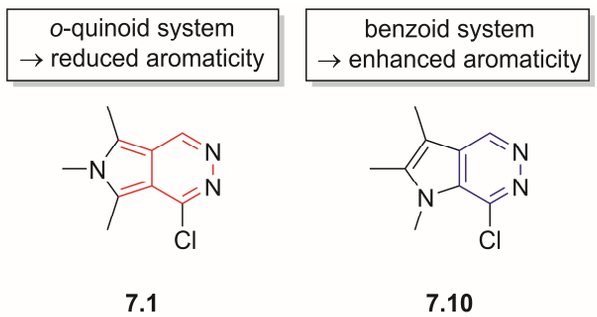


Figure 8.12. The [3′,4′]-pyrrolo annelation reduces the aromaticity of **7.1** compared to the [2′,3′] annelated analog **7.10**.

of the starting fragment is incubated in methanol in the absence of light, suggesting that a different mechanism of activation might be involved.

The reactivity of **7.1** proved to be a lot higher than expected from its structure. It was not flagged as a PAINS compound, nor did a rigorous filter applied before an STD-NMR screening exclude the fragment from the experiment. Therefore, quantum chemical calculations were applied to understand the reactive nature of the compound. Nucleus-independent chemical shift (NICS) calculations indicate a reduced aromaticity for the 3-chloropyridazine ring. The more imine-like nature of the C=N bond is further supported by the observation of an increased bond order. The defining structural feature for these properties is the [3',4']-pyrrolo annelation of the pyridazine. This isoindol-like configuration prevents a continuous delocalization of the π-electron system throughout the molecule, resulting in a less aromatic, and thus, more reactive heterocycle (Figure 8.12). Overall, the high electrophilicity index that is calculated for the molecule, indicates a high reactivity in nucleophilic reactions, which might facilitate the initial step of the reaction cascade. For **7.2**, the postulated product of the reaction cascade, an even higher electrophilicity index is found. This observation might explain why the molecule could not be isolated. Its high reactivity likely results in a quick degradation in solution. Presumably the molecule is only sufficiently stabilized in the environment of the protein pocket.

The potentially very potent inhibitor **7.2** is bound to EP by multiple strong interactions (Figure 7.4). With its aldehyde moiety, the triazole heterocycle form hydrogen bonds to Thr222. Another hydrogen bond to Tyr226 is mediated by a water molecule. These interactions require a significant rearrangement of the surrounding protein residues. Perhaps of even greater importance are the electrostatic interactions evident from the crystal structure. The negatively charged Asp15 side chain is stacked directly above the permanently positively charged pyrrolopyridazine moiety. In addition to that, the least-squares refinement of the crystal structure reveals a curious detail of the interaction. The carbon atom of the formally charged

pyridazinium ion is significantly shifted out of the ring plane. This finding is in line with the hypothesis that a change of the hybridization state, along with a partial charge transfer from the hydroxy group of Thr223, mitigate the positive charge in the heterocycle and lead to a stabilization of the highly reactive molecule.

The biological activity that was detected for **7.1** in multiple assays does not originate from the interactions of the parent molecule alone. This compound was found to react in solution, forming a number of different species that all contribute to the binding affinity of the analyzed mixtures. The most significant contribution to the binding affinity probably emanates from a highly instable trimeric form of the original molecule (**7.2**), which was found to engage EP in strong electrostatic interactions.

Bibliography

1. Copeland, R.A. (2016) The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.*, **15**, 87–95.

2. Ladbury, J.E., Klebe, G., and Freire, E. (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. *Nat. Rev. Drug Discov.*, **9**, 23–27.

3. Dunitz, J.D. (1995) Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.*, **2**, 709–712.

4. Olsson, T.S.G., Ladbury, J.E., Pitt, W.R., and Williams, M.A. (2011) Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein Sci.*, **20**, 1607–1618.

5. Klebe, G. (2015) Applying thermodynamic profiling in lead finding and optimization. *Nat. Rev. Drug Discov.*, **14**, 95–110.

6. Frank, H.S., and Evans, M.W. (1945) Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *J. Chem. Phys.*, **13**, 507–532.

7. Kauzmann, W. (1959) Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.*, **14**, 1–63.

8. Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly. *Nature*, **437**, 640–647.

9. Bingham, R.J., Findlay, J.B.C., Hsieh, S.-Y., et al. (2004) Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein. *J. Am. Chem. Soc.*, **126**, 1675–81.

10. Homans, S.W. (2007) Water, water everywhere - except where it matters? *Drug Discov. Today*, **12**, 534–539.

11. Snyder, P.W., Mecinovic, J., Moustakas, D.T., et al. (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl. Acad. Sci.*, **108**, 17889–17894.

12. Turner, J., Soper, A.K., and Finney, J.L. (1990) A neutron-diffraction study of tetramethylammonium chloride in aqueous solution. *Mol. Phys.*, **70**, 679–700.

13. Buchanan, P., Aldiwan, N., Soper, A.K., et al. (2005) Decreased structure on dissolving methane in water. *Chem. Phys. Lett.*, **415**, 89–93.

14. Snyder, P.W., Lockett, M.R., Moustakas, D.T., and Whitesides, G.M. (2014) Is it the shape of the cavity, or the shape of the water in the cavity? *Eur. Phys. J. Spec. Top.*, **223**, 853–891.

15. Krimmer, S.G., Betz, M., Heine, A., and Klebe, G. (2014) Methyl, Ethyl, Propyl, Butyl: Futile But Not for Water, as the Correlation of Structure and Thermodynamic Signature Shows in a Congeneric Series of Thermolysin Inhibitors. *ChemMedChem*, **9**, 833–846.

16. Krimmer, S.G., Cramer, J., Betz, M., et al. (2016) Rational Design of Thermodynamic and Kinetic Binding Profiles by Optimizing Surface Water Networks Coating Protein-Bound Ligands. *J. Med. Chem.*, **59**, 10530–10548.

17. Wiseman, T., Williston, S., Brandts, J.F., and Lin, L.-N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.*, **179**, 131–137.

18. Copeland, R.A. (2015) The drug–target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.*, **15**, 87–95.

19. Swinney, D. (2006) Biochemical Mechanisms of New Molecular Entities (NMEs) Approved by United States FDA During 2001-2004: Mechanisms Leading to Optimal Efficacy and Safety. *Curr. Top. Med. Chem.*, **6**, 461–478.

20. Copeland, R.A. (2011) Conformational adaptation in drug-target interactions and residence time. *Futur. Med. Chem.*, **3**, 1491–1501.

21. Copeland, R.A., Pompliano, D.L., and Meek, T.D. (2006) Drug–target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.*, **5**, 730–739.

22. Copeland, R.A. (2010) The dynamics of drug-target interactions: drug-target residence time and its impact on efficacy and safety. *Expert Opin. Drug Discov.*, **5**, 305–310.

23. Tummino, P.J., and Copeland, R.A. (2008) Residence Time of Receptor–Ligand Complexes and Its Effect on Biological Function. *Biochemistry*, **47**, 5481–5492.

24. Lu, H., and Tonge, P.J. (2010) Drug–target residence time: critical information for lead optimization. *Curr. Opin. Chem. Biol.*, **14**, 467–474.

25. Lu, H., England, K., am Ende, C., et al. (2009) Slow-Onset Inhibition of the FabI Enoyl Reductase from *Francisella tularensis* : Residence Time and *in Vivo* Activity. *ACS Chem. Biol.*, **4**, 221–231.

26. Guo, D., Mulder-Krieger, T., IJzerman, A.P., and Heitman, L.H. (2012) Functional efficacy of adenosine A2A receptor agonists is positively correlated to their receptor residence time. *Br. J. Pharmacol.*, **166**, 1846–1859.

27. Walkup, G.K., You, Z., Ross, P.L., et al. (2015) Translating slow-binding inhibition kinetics into cellular and in vivo effects. *Nat. Chem. Biol.*, **11**, 416–423.

28. Lundström, I. (1994) Real-time biospecific interaction analysis. *Biosens. Bioelectron.*, **9**, 725–736.

29. Fägerstam, L.G., Frostell-Karlsson, Å., Karlsson, R., et al. (1992) Biospecific interaction analysis using surface plasmon resonance detection applied to kinetic, binding site and concentration analysis. *J. Chromatogr. A*, **597**, 397–410.

30. Fivash, M., Towler, E.M., and Fisher, R.J. (1998) BIAcore for macromolecular interaction. *Curr. Opin. Biotechnol.*, **9**, 97–101.

31. Tajima, M., Urabe, I., Yutani, K., and Okada, H. (1976) Role of Calcium Ions in the Thermostability

of Thermolysin and Bacillus subtilis var. amylosacchariticus Neutral Protease. *Eur. J. Biochem.*, **64**, 243–247.

32. Jongeneel, C.V., Bouvier, J., and Bairoch, A. (1989) A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS Lett.*, **242**, 211–214.

33. Englert, L., Biela, A., Zayed, M., et al. (2010) Displacement of disordered water molecules from hydrophobic pocket creates enthalpic signature: Binding of phosphonamidate to the S1'-pocket of thermolysin. *Biochim. Biophys. Acta, Gen. Subj.*, **1800**, 1192–1202.

34. Matthews, B.W., Jansonius, J.N., Colman, P.M., et al. (1972) Three-dimensional Structure of Thermolysin. *Nat. New Biol.*, **238**, 37–41.

35. Bohacek, R., de Lombaert, S., McMartin, C., et al. (1996) Three-Dimensional Models of ACE and NEP Inhibitors and Their Use in the Design of Potent Dual ACE/NEP Inhibitors. *J. Am. Chem. Soc.*, **118**, 8231–8249.

36. Holland, D.R., Barclay, P.L., Danilewicz, J.C., et al. (1994) Inhibition of thermolysin and neutral endopeptidase 24.11 by a novel glutaramide derivative: X-ray structure determination of the thermolysin-inhibitor complex. *Biochemistry*, **33**, 51–56.

37. Tiraboschi, G., Jullian, N., Thery, V., et al. (1999) A three-dimensional construction of the active site (region 507-749) of human neutral endopeptidase (EC.3.4.24.11). *Protein Eng., Des. Sel.*, **12**, 141–149.

38. Suda, H., Aoyagi, T., Takeuchi, T., and Umezawa, H. (1973) A Thermolysin Inhibitor Produced by Actinomycetes: Phosphoramidon. *J. Antibiot. (Tokyo).*, **26**, 621–623.

39. Bartlett, P.A., and Marlowe, C.K. (1983) Phosphonamidates as transition-state analog inhibitors of thermolysin. *Biochemistry*, **22**, 4618–4624.

40. Jacobsen, N.E., and Bartlett, P.A. (1981) A phosphonamidate dipeptide analog as an inhibitor of carboxypeptidase A. *J. Am. Chem. Soc.*, **103**, 654–657.

41. Elliott, R.L., Marks, N., Berg, M.J., and Portoghese, P.S. (1985) Synthesis and biological evaluation of phosphonamidate peptide inhibitors of enkephalinase and angiotensin-converting enzyme. *J. Med. Chem.*, **28**, 1208–1216.

42. Bertenshaw, S.R., Rogers, R.S., Stern, M.K., et al. (1993) Phosphorus-containing inhibitors of endothelin converting enzyme: Effects of the electronic nature of phosphorus on inhibitor potency. *J. Med. Chem.*, **36**, 173–176.

43. Kemnitz, C.R., and Loewen, M.J. (2007) "Amide resonance" correlates with a breadth of C-N rotation barriers. *J. Am. Chem. Soc.*, **129**, 2521–2528.

44. Bordwell, F.G., Bartmess, J.E., and Hautala, J.A. (1978) Alkyl effects on equilibrium acidities of carbon acids in protic and dipolar aprotic media and the gas phase. *J. Org. Chem.*, **43**, 3095–3101.

45. Bordwell, F.G., Fried, H.E., Hughes, D.L., et al. (1990) Acidities of carboxamides, hydroxamic acids, carbohydrazides, benzenesulfonamides, and benzenesulfonohydrazides in DMSO solution. *J. Org. Chem.*, **55**, 3330–3336.

46. Rahil, J., and Haake, P. (1981) Reactivity and mechanism of hydrolysis of phosphonamides. *J. Am.*

Chem. Soc., **103**, 1723–1734.

47. Hanson, J.E., Kaplan, A.P., and Bartlett, P.A. (1989) Phosphonate analogues of carboxypeptidase A substrates are potent transition-state analogue inhibitors. *Biochemistry*, **28**, 6294–305.

48. Yamauchi, K., Ohtsuki, S., and Kinoshita, M. (1984) Synthesis of peptide analogs containing (2-aminoethyl)phosphonic acid (ciliatine). *J. Org. Chem.*, **49**, 1158–1163.

49. Hariharan, M., Chaberek, S., and Martell, A.E. (1973) Synthesis of Phosphonopeptide Derivatives. *Synth. Commun.*, **3**, 375–379.

50. Hirschmann, R., Yager, K.M., Taylor, C.M., et al. (1997) Phosphonate Diester and Phosphonamide Synthesis. Reaction Coordinate Analysis by 31P NMR Spectroscopy: Identification of Pyrophosphonate Anhydrides and Highly Reactive Phosphonylammonium Salts. *J. Am. Chem. Soc.*, **119**, 8177–8190.

51. Musiol, H.-J., Grams, F., Rudolph-Boehner, S., and Moroder, L. (1994) On the Synthesis of Phosphonamidate Peptides. *J. Org. Chem.*, **59**, 6144–6146.

52. Le Corre, S.S., Berchel, M., Couthon-Gourvès, H., et al. (2014) Atherton–Todd reaction: mechanism, scope and applications. *Beilstein J. Org. Chem.*, **10**, 1166–1196.

53. Burger, A., and Anderson, J.J. (1957) Monoesters and Ester-amidates of Aromatic Phosphonic Acids. *J. Am. Chem. Soc.*, **79**, 3575–3579.

54. Campagne, J.-M., Coste, J., and Jouin, P. (1995) (1H-Benzotriazol-1-yloxy)tris(dimethylamino)-phosphonium Hexafluorophosphate- and (1H-Benzotriazol-1-yloxy)tripyrrolidinophosphonium Hexafluorophosphate-Mediated Activation of Monophosphonate Esters: Synthesis of Mixed Phosphonate Diesters, the Reactivity. *J. Org. Chem.*, **60**, 5214–5223.

55. Ishibashi, Y., Kitamura, M., Kafarski, P., et al. (2009) Solid-phase synthesis of protected α-amino phosphonic acid oligomers. *Chem. Commun.*, **63**, 6985.

56. Fredriksen, K.A., and Amedjkouh, M. (2016) Investigation of Reactive Intermediates and Reaction Pathways in the Coupling Agent Mediated Phos-phonamidation Reaction. *European J. Org. Chem.*, **2016**, 474–482.

57. Erlanson, D.A., McDowell, R.S., and O'Brien, T. (2004) Fragment-based drug discovery. *J. Med. Chem.*, **47**, 3462–3482.

58. Hajduk, P.J., and Greer, J. (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.*, **6**, 211–219.

59. Keserű, G.M., Erlanson, D.A., Ferenczy, G.G., et al. (2016) Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *J. Med. Chem.*, **59**, 8189–8206.

60. Boyd, S.M., Turnbull, A.P., and Walse, B. (2012) Fragment library design considerations. *WIREs Comput. Mol. Sci.*, **2**, 868–885.

61. Congreve, M., Carr, R., Murray, C., and Jhoti, H. (2003) A "Rule of Three" for fragment-based lead discovery? *Drug Discov. Today*, **8**, 876–877.

62. Scott, D.E., Coyne, A.G., Hudson, S.A., and Abell, C. (2012) Fragment-Based Approaches in Drug

Discovery and Chemical Biology. *Biochemistry*, **51**, 4990–5003.

63. Silvestre, H.L., Blundell, T.L., Abell, C., and Ciulli, A. (2013) Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery. *Proc. Natl. Acad. Sci.*, **110**, 12984–12989.

64. Schiebel, J., Radeva, N., Krimmer, S.G., et al. (2016) Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study. *ACS Chem. Biol.*, **11**, 1693–1701.

65. Schiebel, J., Radeva, N., Köster, H., et al. (2015) One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem*, **10**, 1511–1521.

66. Radeva, N., Schiebel, J., Wang, X., et al. (2016) Active Site Mapping of an Aspartic Protease by Multiple Fragment Crystal Structures: Versatile Warheads To Address a Catalytic Dyad. *J. Med. Chem.*, **59**, 9743–9759.

67. Radeva, N., Krimmer, S.G., Stieler, M., et al. (2016) Experimental Active-Site Mapping by Fragments: Hot Spots Remote from the Catalytic Center of Endothiapepsin. *J. Med. Chem.*, **59**, 7561–7575.

68. Schiebel, J., Krimmer, S.G., Rower, K., et al. (2016) High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure*, **24**, 1398–1409.

69. Diller, D.J., and Hobbs, D.W. (2004) Deriving knowledge through data mining high-throughput screening data. *J. Med. Chem.*, **47**, 6373–6383.

70. Schorpp, K., Rothenaigner, I., Salmina, E., et al. (2014) Identification of Small-Molecule Frequent Hitters from AlphaScreen High-Throughput Screens. *J. Biomol. Screen.*, **19**, 715–726.

71. Soares, K.M., Blackmon, N., Shun, T.Y., et al. (2010) Profiling the NIH Small Molecule Repository for compounds that generate H2O2 by redox cycling in reducing environments. *Assay Drug Dev. Technol.*, **8**, 152–74.

72. Huth, J.R., Mendoza, R., Olejniczak, E.T., et al. (2005) ALARM NMR: A rapid and robust experimental method to detect reactive false positives in biochemical screens. *J. Am. Chem. Soc.*, **127**, 217–224.

73. Dahlin, J.L., Nissink, J.W.M., Strasser, J.M., et al. (2015) PAINS in the Assay: Chemical Mechanisms of Assay Interference and Promiscuous Enzymatic Inhibition Observed during a Sulfhydryl-Scavenging HTS. *J. Med. Chem.*, **58**, 2091–2113.

74. Feng, B.Y., Simeonov, A., Jadhav, A., et al. (2007) A high-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.*, **50**, 2385–2390.

75. Tarzia, G., Antonietti, F., Duranti, A., et al. (2007) Identification of a Bioactive Impurity in a Commercial Sample of 6-Methyl-2-p-Tolylaminobenzo[d][1,3]Oxazin-4-One (URB754). *Ann. Chim.*, **97**, 887–894.

76. Hermann, J.C., Chen, Y., Wartchow, C., et al. (2013) Metal Impurities Cause False Positives in High-Throughput Screening Campaigns. *ACS Med. Chem. Lett.*, **4**, 197–200.

77. Jaki, B.U., Franzblau, S.G., Chadwick, L.R., et al. (2008) Purity–Activity Relationships of Natural

Products: The Case of Anti-TB Active Ursolic Acid. *J. Nat. Prod.*, **71**, 1742–1748.

78. Devine, S.M., Mulcair, M.D., Debono, C.O., et al. (2015) Promiscuous 2-Aminothiazoles (PrATs): A Frequent Hitting Scaffold. *J. Med. Chem.*, **58**, 1205–1214.

79. Schneider, P., Röthlisberger, M., Reker, D., et al. (2016) Spotting and designing promiscuous ligands for drug discovery. *Chem. Commun.*, **52**, 1135–1138.

80. Simeonov, A., Jadhav, A., Thomas, C.J., et al. (2008) Fluorescence Spectroscopic Profiling of Compound Libraries. *J. Med. Chem.*, **51**, 2363–2371.

81. Baell, J.B., and Holloway, G.A. (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.*, **53**, 2719–2740.

82. McLeod, D.A., Brinkworth, R.I., Ashley, J.A., et al. (1991) Phosphonamidates and phosphonamidate esters as HIV-1 protease inhibitors. *Bioorg. Med. Chem. Lett.*, **1**, 653–658.

83. Mucha, A., Kunert, A., Grembecka, J., et al. (2006) A phosphonamidate containing aromatic N-terminal amino group as inhibitor of leucine aminopeptidase-design, synthesis and stability. *Eur. J. Med. Chem.*, **41**, 768–772.

84. Mookhtiar, K.A., Marlowe, C.K., Bartlett, P.A., and van Wart, H.E. (1987) Phosphonamidate inhibitors of human neutrophil collagenase. *Biochemistry*, **26**, 1962–1965.

85. Hua, T.D., Lamaty, F., Souriau, C., et al. (1996) Specific recognition of a tetrahedral phosphonamidate transition state analogue group by a recombinant antibody Fab fragment. *Amino Acids*, **10**, 167–172.

86. Vallee, M.R.J., Majkut, P., Wilkening, I., et al. (2011) Staudinger-phosphonite reactions for the chemoselective transformation of azido-containing peptides and proteins. *Org. Lett.*, **13**, 5440–5443.

87. Vallee, M.R.J., Artner, L.M., Dernedde, J., and Hackenberger, C.P.R. (2013) Alkyne phosphonites for sequential azide-azide couplings. *Angew. Chem., Int. Ed. Engl.*, **52**, 9504–9508.

88. Vallee, M.R.J., Majkut, P., Krause, D., et al. (2015) Chemoselective bioconjugation of triazole phosphonites in aqueous media. *Chem. Eur. J.*, **21**, 970–974.

89. de Medina, P., Ingrassia, L.S., and Mulliez, M.E. (2003) Synthesis of the first stable phosphonamide transition state analogue. *J. Org. Chem.*, **68**, 8424–8430.

90. Mucha, A., Grembecka, J., Cierpicki, T., and Kafarski, P. (2003) Hydrolysis of the Phosphonamidate Bond in Phosphono Dipeptide Analogues--- the Influence of the Nature of the N-Terminal Functional Group. *European J. Org. Chem.*, **2003**, 4797–4803.

91. Christianson, D.W., and Lipscomb, W.N. (1986) Structure of the complex between an unexpectedly hydrolyzed phosphonamidate inhibitor and carboxypeptidase A. *J. Am. Chem. Soc.*, **108**, 545–546.

92. Biela, A., Sielaff, F., Terwesten, F., et al. (2012) Ligand binding stepwise disrupts water network in thrombin: enthalpic and entropic changes reveal classical hydrophobic effect. *J. Med. Chem.*, **55**, 6094–6110.

93. Biela, A., Nasief, N.N., Betz, M., et al. (2013) Dissecting the Hydrophobic Effect on the Molecular Level: The Role of Water, Enthalpy, and Entropy in Ligand Binding to Thermolysin. *Angew. Chemie*

Int. Ed., **52**, 1822–1828.

94. Biela, A., Betz, M., Heine, A., and Klebe, G. (2012) Water Makes the Difference: Rearrangement of Water Solvation Layer Triggers Non-additivity of Functional Group Contributions in Protein-Ligand Binding. *ChemMedChem*, **7**, 1423–1434.

95. Biela, A., Nasief, N.N., Betz, M., et al. (2013) Zerlegung des hydrophoben Effekts auf molekularer Ebene: Die Rolle von Wasser, Enthalpie und Entropie bei der Ligandenbindung an Thermolysin. *Angew. Chemie*, **125**, 1868–1876.

96. Krimmer, S.G., and Klebe, G. (2015) Thermodynamics of protein-ligand interactions as a reference for computational analysis: how to assess accuracy, reliability and relevance of experimental data. *J. Comput. Mol. Des.*, **29**, 867–883.

97. Gruner, S., Neeb, M., Barandun, L.J., et al. (2014) Impact of protein and ligand impurities on ITC-derived protein-ligand thermodynamics. *Biochim. Biophys. Acta*, **1840**, 2843–2850.

98. Kishore Kumar, G.D., Saenz, D., Lokesh, G.L., and Natarajan, A. (2006) Microwave-assisted cleavage of phosphate, phosphonate and phosphoramide esters. *Tetrahedron Lett.*, **47**, 6281–6284.

99. Wilkening, I., del Signore, G., and Hackenberger, C.P.R. (2011) Synthesis of phosphonamidate peptides by Staudinger reactions of silylated phosphinic acids and esters. *Chem. Commun. (Camb).*, **47**, 349–351.

100. Blackburn, G. M., Wentworth, P. Catalytic Antibodies for Carbamate Activation by a non-spontaneous Reaction Mechanism.

101. Hennion, M.-C. (1999) Solid-phase extraction: Method development, sorbents, and coupling with liquid chromatography. *J. Chromatogr. A*, **856**, 3–54.

102. Kamysz, W., Okroj, M., Lempicka, E., et al. (2004) Fast and efficient purification of synthetic peptides by solid-phase extraction. *Acta Chromatogr.*, **14**, 180–186.

103. Nilsson, U.J., Fournier, E.J., and Hindsgaul, O. (1998) Solid-phase extraction on C18 silica as a purification strategy in the solution synthesis of a 1-thio-beta-D-galactopyranoside library. *Bioorg. Med. Chem.*, **6**, 1563–1575.

104. Asada, T., Koi, Y., Arakawa, R., et al. (2014) Isolation techniques for anthocyanidin 3,5-diglucosides and their related chemicals using supramolecules technique, and two solid-phase extraction cartridges. *J. Chromatogr. A*, **1351**, 21–28.

105. Morgan, B., Scholtz, J.M., Ballinger, M.D., et al. (1991) Differential binding energy: A detailed evaluation of the influence of hydrogen-bonding and hydrophobic groups on the inhibition of thermolysin by phosphorus-containing inhibitors. *J. Am. Chem. Soc.*, **113**, 297–307.

106. Pauli, G.F., Chen, S.-N., Simmler, C., et al. (2014) Importance of purity evaluation and the potential of quantitative (1)H NMR as a purity assay. *J. Med. Chem.*, **57**, 9220–9231.

107. Freire, E. (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov. Today*, **13**, 869–874.

108. Ferenczy, G.G., and Keseru, G.M. (2010) Thermodynamics guided lead discovery and optimization. *Drug Discov. Today*, **15**, 919–932.

109. Edink, E., Jansen, C., Leurs, R., and De Esch, I.J.P. (2010) The heat is on: Thermodynamic analysis in fragment-based drug discovery. *Drug Discov. Today Technol.*, **7**, 189–201.

110. Freire, E. (2009) A thermodynamic approach to the affinity optimization of drug candidates. *Chem. Biol. Drug Des.*, **74**, 468–472.

111. Leavitt, S., and Freire, E. (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.*, **11**, 560–566.

112. Holdgate, G.A., and Ward, W.H.J. (2005) Measurements of binding thermodynamics in drug discovery. *Drug Discov. Today*, **10**, 1543–1550.

113. Geschwindner, S., Ulander, J., and Johansson, P. (2015) Ligand binding thermodynamics in drug discovery: still a hot tip? *J. Med. Chem.*, **58**, 6321–6335.

114. Martin, S.F., and Clements, J.H. (2013) Correlating structure and energetics in protein-ligand interactions: paradigms and paradoxes. *Annu. Rev. Biochem.*, **82**, 267–93.

115. Pan, A.C., Borhani, D.W., Dror, R.O., and Shaw, D.E. (2013) Molecular determinants of drug-receptor binding kinetics. *Drug Discov. Today*, **18**, 667–673.

116. Chaires, J.B. (2008) Calorimetry and thermodynamics in drug design. *Annu. Rev. Biophys.*, **37**, 135–151.

117. Whitesides, G.M., and Krishnamurthy, V.M. (2005) Designing ligands to bind proteins. *Q. Rev. Biophys.*, **38**, 385–395.

118. Ball, P. (2008) Water as an active constituent in cell biology. *Chem. Rev.*, **108**, 74–108.

119. Ladbury, J.E. (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.*, **3**, 973–980.

120. Olsson, T.S.G., Williams, M.A., Pitt, W.R., and Ladbury, J.E. (2008) The Thermodynamics of Protein–Ligand Interaction and Solvation: Insights for Ligand Design. *J. Mol. Biol.*, **384**, 1002–1017.

121. Bissantz, C., Kuhn, B., and Stahl, M. (2010) A medicinal chemist's guide to molecular interactions. *J. Med. Chem.*, **53**, 5061–5084.

122. Southall, N.T., Dill, K.A., and Haymet, A.D.J. (2002) A View of the Hydrophobic Effect. *J. Phys. Chem.*, **106**, 521–533.

123. Jelesarov, I., and Bosshard, H.R. (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to invesitigate the energetics of biomolecular recognition. *J. Mol. Recognit.*, **12**, 3–18.

124. Setny, P., Baron, R., and McCammon, J.A. (2010) How Can Hydrophobic Association Be Enthalpy Driven? *J Chem Theory Comput*, **6**, 2866–2871.

125. Hummer, G. (2010) Molecular binding: Under water's influence. *Nat. Chem.*, **2**, 906–7.

126. Breiten, B., Lockett, M.R., Sherman, W., et al. (2013) Water Networks Contribute to Enthalpy / Entropy Compensation in Protein-Ligand Binding. *J. Am. Chem. Soc.*, **135**, 15579–15584.

127. Adekoya, O.A., and Sylte, I. (2009) The thermolysin family (M4) of enzymes: Therapeutic and

biotechnological potential. *Chem. Biol. Drug Des.*, **73**, 7–16.

128. Eijsink, V.G.H., Matthews, B.W., and Vriend, G. (2011) The role of calcium ions in the stability and instability of a thermolysin-like protease. *Protein Sci.*, **20**, 1346–1355.

129. Petrillo, E.W.J., and Ondetti, M.A. (1982) Angiotensin-converting enzyme inhibitors: medicinal chemistry and biological actions. *Med. Res. Rev.*, **2**, 1–41.

130. Gallicchio, E., Kubo, M.M., and Levy, R.M. (1998) Entropy-Enthalpy Compensation in Solvation and Ligand Binding Revisited. *J. Am. Chem. Soc.*, **120**, 4526–4527.

131. Chodera, J.D., and Mobley, D.L. (2013) Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys.*, **42**, 121–42.

132. Olsson, T.S.G., Williams, M.A., Pitt, W.R., and Ladbury, J.E. (2008) The Thermodynamics of Protein-Ligand Interaction and Solvation: Insights for Ligand Design. *J. Mol. Biol.*, **384**, 1002–1017.

133. Betz, M., Wulsdorf, T., Krimmer, S.G., and Klebe, G. (2015) Impact of Surface Water Layers on Protein-Ligand Binding: How Well Are Experimental Data Reproduced by Molecular Dynamics Simulations in a Thermolysin Test Case. *J. Chem. Inf. Model.*, **56**, 223–233.

134. Guo, J., Huang, W., and Scanlan, T.S. (1994) Kinetic and Mechanistic Characterization of an Efficient Hydrolytic Antibody: Evidence for the Formation of an Acyl Intermediate. *J. Am. Chem. Soc.*, **116**, 6062–6069.

135. Holden, H.M., Tronrud, D.E., Monzingo, A.F., et al. (1987) Slow- and fast-binding inhibitors of thermolysin display different modes of binding: crystallographic analysis of extended phosphonamidate transition-state analogues. *Biochemistry*, **26**, 8542–8553.

136. Neudert, G., and Klebe, G. (2011) fconv: format conversion, manipulation and feature computation of molecular data. *Bioinformatics*, **27**, 1021–1022.

137. Winn, M.D., Ballard, C.C., Cowtan, K.D., et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**, 235–242.

138. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

139. Matthews, B.W. (1972) The γ Turn. Evidence for a New Folded Conformation in Proteins. *Macromolecules*, **5**, 818–819.

140. *G. J. Kleywegt, J.-Y. Zou, M. Kjeldgaard, T. A. Jones, in International Tables for Crystallography, Vol. F (Eds.: M. G. Rossmann, E. Arnold), Springer Netherlands, Dordrecht, 2001, 353–356.*

141. Carugo, O. (1999) Correlation between occupancy and B factor of water molecules in protein crystal structures. *Protein Eng.*, **12**, 1021–1024.

142. Jeffrey, G.A. (1997) *An Introduction to Hydrogen Bonding*, Oxford University Press, Oxford.

143. Grabowski, S.J. (2006) *Hydrogen Bonding-New Insights*, Springer, Heidelberg.

144. Roden, L.D., and Myszka, D.G. (1996) Global Analysis of a Macromolecular Interaction Measured on BIAcore. *Biochem. Biophys. Res. Commun.*, **225**, 1073–1077.

145. Karlsson, R., and Fält, A. (1997) Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors. *J. Immunol. Methods*, **200**, 121–133.

146. Carugo, O., and Bordo, D. (1999) How many water molecules can be detected by protein crystallography? *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **55**, 479–483.

147. Fenley, A.T., Muddana, H.S., and Gilson, M.K. (2012) Entropy-enthalpy transduction caused by conformational shifts can obscure the forces driving protein-ligand binding. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 20006–200011.

148. Vaitheeswaran, S., Yin, H., Rasaiah, J.C., and Hummer, G. (2004) Water clusters in nonpolar cavities. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 17002–17005.

149. Lee, J., and Kim, S.H. (2009) Water polygons in high-resolution protein crystal structures. *Protein Sci.*, **18**, 1370–1376.

150. Winquist, J., Geschwindner, S., Xue, Y., et al. (2013) Identification of Structural–Kinetic and Structural–Thermodynamic Relationships for Thrombin Inhibitors. *Biochemistry*, **52**, 613–626.

151. Chemical Computing Group Inc. Molecular Operating Environment (MOE), Chemical Computing Group Inc., Montreal, QC, Canada.

152. Bayly, C.I., Cieplak, P., Cornell, W.D., and Kollman, P.A. (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.*, **97**, 10269–10280.

153. Frisch, M.J., Trucks, G.W., Schlegel, H.B., et al. Gaussian09, Revision D.01 (2009), Gaussian Inc., Wallingford, CT.

154. Case, D.A., Babin, V., Berryman, J.T., et al. AMBER 14 (2014), University of California, San Francisco.

155. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *J. Mol. Graph.*, **14**, 33–38.

156. Holmes, M.A., and Matthews, B.W. (1982) Structure of thermolysin refined at 1.6 Å resolution. *J. Mol. Biol.*, **160**, 623–639.

157. Mueller, U., Darowski, N., Fuchs, M.R., et al. (2012) Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J. Synchrotron Radiat.*, **19**, 442–449.

158. Kabsch, W. (2010) XDS. *Acta Crystallogr., Sect. D Biol. Crystallogr.*, **66**, 125–132.

159. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., et al. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.

160. Holland, D.R., Tronrud, D.E., Pley, H.W., et al. (1992) Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry*, **31**, 11310–11316.

161. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 486–501.

162. Adams, P.D., Afonine, P. V, Bunkoczi, G., et al. (2010) PHENIX: a comprehensive Python-based

system for macromolecular structure solution. *Acta Crystallogr., Sect. D Biol. Crystallogr.*, **66**, 213–221.

163. Moriarty, N.W., Grosse-Kunstleve, R.W., and Adams, P.D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr., Sect. D Biol. Crystallogr.*, **65**, 1074–1080.

164. Schrödinger LLC The PyMOL Molecular Graphics System, Schrödinger LLC, New York, NY (USA).

165. Tellinghuisen, J. (2012) Designing isothermal titration calorimetry experiments for the study of 1:1 binding: problems with the "standard protocol." *Anal. Biochem.*, **424**, 211–220.

166. Mizoue, L.S., and Tellinghuisen, J. (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. *Anal. Biochem.*, **326**, 125–127.

167. Keller, S., Vargas, C., Zhao, H., et al. (2012) High-precision isothermal titration calorimetry with automated peak shape analysis. *Anal. Chem.*, **84**, 5066–5073.

168. Houtman, J.C.D., Brown, P.H., Bowden, B., et al. (2007) Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. *Protein Sci.*, **16**, 30–42.

169. Brautigam, C.A. (2015) Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Methods Enzymol.*, **562**, 109–133.

170. Krissinel, E., and Henrick, K. (2007) Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.*, **372**, 774–797.

171. Bernstein, F.C., Koetzle, T.F., Williams, G.J., et al. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem.*, **80**, 319–324.

172. Falconer, R.J. (2016) Applications of isothermal titration calorimetry - the research and technical developments from 2011 to 2015. *J. Mol. Recognit.*, **29**, 504–515.

173. Lim, V.I., Curran, J.F., and Garber, M.B. (2012) Hydration shells of molecules in molecular association: A mechanism for biomolecular recognition. *J. Theor. Biol.*, **301**, 42–8.

174. Mondal, J., Friesner, R.A., and Berne, B.J. (2014) Role of Desolvation in Thermodynamics and Kinetics of Ligand Binding to a Kinase. *J. Chem. Theory Comput.*, **10**, 5696–5705.

175. Berne, B.J., Weeks, J.D., and Zhou, R. (2009) Dewetting and Hydrophobic Interaction in Physical and Biological Systems. *Annu. Rev. Phys. Chem.*, **60**, 85–103.

176. Snyder, P.W., Mecinovic, J., Moustakas, D.T., et al. (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 17889–94.

177. Haider, K., Wickstrom, L., Ramsey, S., et al. (2016) Enthalpic Breakdown of Water Structure on Protein Active-Site Surfaces. *J. Phys. Chem. B*, **120**, 8743–8756.

178. Andrade, J.M., and Estévez-Pérez, M.G. (2014) Statistical comparison of the slopes of two regression lines: A tutorial. *Anal. Chim. Acta*, **838**, 1–12.

179. Cramer, J., Krimmer, S.G., Fridh, V., et al. (2017) Elucidating the Origin of Long Residence Time

Binding for Inhibitors of the Metalloprotease Thermolysin. *ACS Chem. Biol.*, **12**, 225–233.

180. CCP4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **50**, 760–763.

181. Matthews, B.W. (1972) The g Turn. Evidence for a New Folded Conformation in Proteins. *Macromolecules*, **5**, 818–819.

182. Kleywegt, G.J. MOLEMAN - unpublished program.

183. Cramer, J., and Klebe, G. (2017) An Allyl Protection and Improved Purification Strategy Enables the Synthesis of Functionalized Phosphonamidate Peptides. *Synthesis (Stuttg).*, **49**, DOI: 10.1055/s-0036-1588393.

184. Piszczek, G. (2015) SEDPHAT – A platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods*, **76**, 137–148.

185. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr., Sect. D Biol. Crystallogr.*, **66**, 133–144.

186. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **60**, 2126–2132.

187. Burling, F.T., Weis, W.I., Flaherty, K.M., and Brünger, A.T. (1996) Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science*, **271**, 72–77.

188. Kleywegt, G.J., and Zou, J.-Y. (2001) Around O, in *International Tables for Crystallography Volume F: Crystallography of biological macromolecules* (eds.Rossmann, M.G., and Arnold, E.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 353–356.

189. Schiltz, M., Fourme, R., and Prangé, T. (2003) Use of Noble Gases Xenon and Krypton as Heavy Atoms in Protein Structure Determination. *Methods Enzymol.*, **374**, 83–119.

190. Levitt, M., and Park, B.H. (1993) Water: Now you see it, now you don't. *Structure*, **1**, 223–226.

191. Read, R.J. (2001) Model phases: probabilities, bias and maps, in *International Tables for Crystallography Volume F: Crystallography of biological macromolecules* (eds.Rossmann, M.G., and Arnold, E.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 325–331.

192. Brünger, A.T., and Rice, L.M. (1997) Crystallographic refinement by simulated annealing: Methods and applications. *Methods Enzymol.*, **277**, 243–269.

193. Hodel, A., Kim, S.-H., and Brünger, A.T. (1992) Model bias in macromolecular crystal structures. *Acta Cryst. A*, **48**, 851–858.

194. Quillin, M.L., Wingfield, P.T., and Matthews, B.W. (2006) Determination of solvent content in cavities in IL-1beta using experimentally phased electron density. *Proc. Natl. Acad. Sci. USA*, **103**, 19749–19753.

195. Liu, L., Quillin, M.L., and Matthews, B.W. (2008) Use of experimental crystallographic phases to examine the hydration of polar and nonpolar cavities in T4 lysozyme. *Proc. Natl. Acad. Sci. USA*, **105**, 14406–14411.

196. Urzhumtsev, A., Afonine, P.V., Lunin, V.Y., et al. (2014) Metrics for comparison of crystallographic

maps. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **70**, 2593–2606.

197. Ernst, J.A., Clubb, R.T., Zhou, H.X., et al. (1995) Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science*, **267**, 1813–1817.

198. Yu, B., Blaber, M., Gronenborn, A.M., et al. (1999) Disordered water within a hydrophobic protein cavity visualized by x-ray crystallography. *Proc. Natl. Acad. Sci. USA*, **96**, 103–108.

199. Matthews, B.W., Morton, A.G., and Dahlquist, F.W. (1995) Use of NMR to Detect Water Within Nonpolar Protein Cavities. *Science*, **270**, 1847–1849.

200. Sonavane, S., and Chakrabarti, P. (2008) Cavities and Atomic Packing in Protein Structures and Interfaces. *PLoS Comput. Biol.*, **4**, e1000188.

201. Hubbard, S.J., Gross, K.H., and Argos, P. (1994) Intramolecular Cavities in Globular Proteins. *Protein Eng.*, **7**, 613–626.

202. Wolfenden, R., and Radzicka, A. (1994) On the probability of finding a water molecule in a nonpolar cavity. *Science*, **265**, 936–937.

203. Bhat, S., and Purisima, E.O. (2006) Molecular surface generation using a variable-radius solvent probe. *Proteins Struct. Funct. Genet.*, **62**, 244–261.

204. Li, A.J.-J., and Nussinov, R. (1998) A set of van der Waals and Coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins Struct. Funct. Genet.*, **32**, 111–127.

205. English, A.C., Groom, C.R., and Hubbard, R.E. (2001) Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng.*, **14**, 47–59.

206. Prangé, T., Schiltz, M., Pernot, L., et al. (1998) Exploring hydrophobic sites in proteins with xenon or krypton. *Proteins Struct. Funct. Genet.*, **30**, 61–73.

207. Tanwar, A.S., Goyal, V.D., Choudhary, D., et al. (2013) Importance of Hydrophobic Cavities in Allosteric Regulation of Formylglycinamide Synthetase: Insight from Xenon Trapping and Statistical Coupling Analysis. *PLoS One*, **8**, e77781.

208. Quillin, M.L., Breyer, W.A., Griswold, I.J., and Matthews, B.W. (2000) Size versus polarizability in protein-ligand interactions: binding of noble gases within engineered cavities in phage T4 lysozyme. *J. Mol. Biol.*, **302**, 955–977.

209. Schlitz, M., Shepard, W., Fourme, R., et al. (1997) High-Pressure Krypton Gas and Statistical Heavy-Atom Refinement: a Successful Combination of Tools for Macromolecular Structure Determination. *Acta Crystallogr D Biol Crystallogr*, **53**, 78–92.

210. Matthews, B.W., and Liu, L. (2009) A review about nothing: Are apolar cavities in proteins really empty? *Protein Sci.*, **18**, 494–502.

211. Adamek, D.H., Guerrero, L., Blaber, M., and Caspar, D.L.D. (2005) Structural and energetic consequences of mutations in a solvated hydrophobic cavity. *J. Mol. Biol.*, **346**, 307–318.

212. Xu, J., Baase, W.A., Quillin, M.L., et al. (2001) Structural and thermodynamic analysis of the binding of solvent at internal sites in T4 lysozyme. *Protein Sci.*, **10**, 1067–1078.

213. Baker, B.M., and Murphy, K.P. (1996) Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. *Biophys. J.*, **71**, 2049–2055.

214. Leung, C.S., Leung, S.S.F., Tirado-Rives, J., and Jorgensen, W.L. (2012) Methyl Effects on Protein–Ligand Binding. *J. Med. Chem.*, **55**, 4489–4500.

215. Young, T., Abel, R., Kim, B., et al. (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. USA*, **104**, 808–813.

216. Wang, L., Berne, B.J., and Friesner, R.A. (2011) Ligand binding to protein-binding pockets with wet and dry regions. *Proc. Natl. Acad. Sci. USA*, **108**, 1326–1330.

217. Barratt, E., Bingham, R.J., Warner, D.J., et al. (2005) Van der Waals Interactions Dominate Ligand-Protein Association in a Protein Binding Site Occluded from Solvent Water. *J. Am. Chem. Soc.*, **127**, 11827–11834.

218. Lee, C., Maeng, J.S., Kocher, J.P., et al. (2001) Cavities of alpha(1)-antitrypsin that play structural and functional roles. *Protein Sci.*, **10**, 1446–1453.

219. Vallone, B., and Brunori, M. (2004) Roles for holes: are cavities in proteins mere packing defects? *Ital. J. Biochem.*, **53**, 46–53.

220. Schiebel, J., Gaspari, R., Sandner, A., et al. (2017) Charges Shift Protonation: Ultimately, Neutron Diffraction Discloses that Aniline and 2-Aminopyridine Become Protonated Upon Binding to Trypsin. *Angew. Chemie - Int. Ed.*

221. Eriksson, A.E., Baase, W.A., Zhang, X.-J., et al. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science*, **255**, 178–183.

222. Kellis, J.T., Nyberg, K., and Fersht, A.R. (1989) Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry*, **28**, 4914–4922.

223. Oxford Cryosystems (1999) Portable xenon pressure chamber. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **55**, 724.

224. Mueller, U., Förster, R., Hellmig, M., et al. (2015) The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. *Eur. Phys. J. Plus*, **130**, 141.

225. Thorn, A., and Sheldrick, G.M. (2011) ANODE: Anomalous and heavy-atom density calculation. *J. Appl. Crystallogr.*, **44**, 1285–1287.

226. Schneider, T.R., and Sheldrick, G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58**, 1772–1779.

227. Sheldrick, G.M. (2007) A short history of SHELX. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **64**, 112–122.

228. Pape, T., and Schneider, T.R. (2004) HKL2MAP : a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.*, **37**, 843–844.

229. Cowtan, K. (1994) "dm": an automated procedure for phase improvement by density modification. *Jt. CCP4 ESF-EACBM Newsl. Protein Crystallogr.*, **31**, 34–38.

230. Lamzin, V.S., and Wilson, K.S. (1993) Automated refinement of protein models. *Acta Crystallogr D Biol Crystallogr*, **49**, 129–147.

231. Durrant, J.D., Votapka, L., Amaro, R.E., et al. (2014) POVME 2.0: An enhanced tool for determining pocket shape and volume characteristics. *J. Chem. Theory Comput.*, **10**, 5047–5056.

232. Qvist, J., Davidovic, M., Hamelberg, D., and Halle, B. (2008) A dry ligand-binding cavity in a solvated protein. *Proc. Natl. Acad. Sci. USA*, **105**, 6296–6301.

233. Gerstein, M., Richards, F.M., Chapman, M.S., and Connolly, M.L. (2001) Protein surfaces and volumes: measurement and use, in *International Tables for Crystallography Volume F: Crystallography of biological macromolecules* (eds.Rossmann, M.G., and Arnold, E.), Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 531–545.

234. Voss, N.R., and Gerstein, M. (2010) 3V: Cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.*, **38**, 1–8.

235. Lang, P.T., Holton, J.M., Fraser, J.S., and Alber, T. (2014) Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proc. Natl. Acad. Sci. USA*, **111**, 237–242.

236. Kleywegt, G.J., and Jones, T.A. (1996) xdlMAPMAN and xdlDATAMAN - Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **52**, 826–828.

237. Inouye, K., Kuzuya, K., and Tonomura, B. (1998) Effect of salts on the solubility of thermolysin: a remarkable increase in the solubility as well as the activity by the addition of salts without aggregation or dispersion of thermolysin. *J. Biochem.*, **123**, 847–852.

238. Zhang, Y.-L., and Zhang, Z.-Y. (1998) Low-Affinity Binding Determined by Titration Calorimetry Using a High-Affinity Coupling Ligand: A Thermodynamic Study of Ligand Binding to Protein Tyrosine Phosphatase 1B. *Anal. Biochem.*, **261**, 139–148.

239. Scheuermann, T.H., and Brautigam, C.A. (2015) High-precision, automated integration of multiple isothermal titration calorimetric thermograms: New features of NITPIC. *Methods*, **76**, 87–98.

240. Brautigam, C.A., Zhao, H., Vargas, C., et al. (2016) Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions. *Nat. Protoc.*, **11**, 882–894.

241. Goldberg, R.N., Kishore, N., and Lennen, R.M. (2002) Thermodynamic quantities for the ionization reaction of buffers. *J. Phys. Chem. Ref. Data*, **31**, 231–370.

242. Schoop, A., and Dey, F. (2015) On-rate based optimization of structure - kinetic relationship-surfing the kinetic map. *Drug Discov. today. Technol.*, **17**, 9–15.

243. Guo, D., Heitman, L.H., and IJzerman, A.P. (2015) The Role of Target Binding Kinetics in Drug Discovery. *ChemMedChem*, **10**, 1793–1796.

244. Segala, E., Guo, D., Cheng, R.K.Y., et al. (2016) Controlling the Dissociation of Ligands from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength. *J. Med. Chem.*, **59**, 6470–6479.

245. Chang, A., Schiebel, J., Yu, W., et al. (2013) Rational optimization of drug-target residence time: insights from inhibitor binding to the Staphylococcus aureus FabI enzyme-product complex. *Biochemistry*, **52**, 4217–4228.

246. Miller, D.C., Lunn, G., Jones, P., et al. (2012) Investigation of the effect of molecular properties on the binding kinetics of a ligand to its biological target. *Medchemcomm*, **3**, 449.

247. Bortolato, A., Tehan, B.G., Bodnarchuk, M.S., et al. (2013) Water Network Perturbation in Ligand Binding: Adenosine A 2A Antagonists as a Case Study. *J. Chem. Inf. Model.*, **53**, 1700–1713.

248. Pearlstein, R.A., Sherman, W., and Abel, R. (2013) Contributions of water transfer energy to protein-ligand association and dissociation barriers: Watermap analysis of a series of p38a MAP kinase inhibitors. *Proteins Struct. Funct. Bioinforma.*, **81**, 1509–1526.

249. Schmidtke, P., Luque, F.J., Murray, J.B., and Barril, X. (2011) Shielded Hydrogen Bonds as Structural Determinants of Binding Kinetics: Application in Drug Design. *J. Am. Chem. Soc.*, **133**, 18903–18910.

250. Carroll, M.J., Mauldin, R. V, Gromova, A. V, et al. (2012) Evidence for dynamics in proteins as a mechanism for ligand dissociation. *Nat. Chem. Biol.*, **8**, 246–252.

251. Luckner, S.R., Liu, N., Am Ende, C.W., et al. (2010) A Slow, Tight Binding Inhibitor of InhA, the Enoyl-Acyl Carrier Protein Reductase from Mycobacterium tuberculosis. *J. Biol. Chem.*, **285**, 14330–14337.

252. Kruse, A.C., Hu, J., Pan, A.C., et al. (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature*, **482**, 552–556.

253. Liu, Y., Stoll, V.S., Richardson, P.L., et al. (2004) Hepatitis C NS3 protease inhibition by peptidyl-alpha-ketoamide inhibitors: kinetic mechanism and structure. *Arch. Biochem. Biophys.*, **421**, 207–216.

254. Garvey, E.P., Schwartz, B., Gartland, M.J., et al. (2009) Potent inhibitors of HIV-1 integrase display a two-step, slow-binding inhibition mechanism which is absent in a drug-resistant T66I/M154I mutant. *Biochemistry*, **48**, 1644–1653.

255. Hausrath, A.C., and Matthews, B.W. (2002) Thermolysin in the absence of substrate has an open conformation. *Acta Crystallogr., Sect. D Biol. Crystallogr.*, **58**, 1002–1007.

256. Holland, D.R., Hausrath, A.C., Juers, D., and Matthews, B.W. (1995) Structural analysis of zinc substitutions in the active site of thermolysin. *Protein Sci.*, **4**, 1955–1965.

257. Arolas, J.L., Botelho, T.O., Vilcinskas, A., and Gomis-Ruth, F.X. (2011) Structural evidence for standard-mechanism inhibition in metallopeptidases from a complex poised to resynthesize a peptide bond. *Angew. Chem., Int. Ed. Engl.*, **50**, 10357–10360.

258. Sonnhammer, E.L., Eddy, S.R., and Durbin, R. (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins*, **28**, 405–420.

259. McMurray, J.J. V, Packer, M., Desai, A.S., et al. (2014) Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N. Engl. J. Med.*, **371**, 993–1004.

260. Smollich, M., Gotte, M., Yip, G.W., et al. (2007) On the role of endothelin-converting enzyme-1 (ECE-1) and neprilysin in human breast cancer. *Breast Cancer Res. Treat.*, **106**, 361–369.

261. Funalot, B., Ouimet, T., Claperon, A., et al. (2004) Endothelin-converting enzyme-1 is expressed in human cerebral cortex and protects against Alzheimer's disease. *Mol. Psychiatry*, **9**, 1059, 1122–1128.

262. Yasukawa, K., Kusano, M., and Inouye, K. (2007) A new method for the extracellular production of

recombinant thermolysin by co-expressing the mature sequence and pro-sequence in Escherichia coli. *Protein Eng., Des. Sel.*, **20**, 375–383.

263. Kusano, M., Yasukawa, K., Hashida, Y., and Inouye, K. (2006) Engineering of the pH-dependence of thermolysin activity as examined by site-directed mutagenesis of Asn112 located at the active site of thermolysin. *J. Biochem.*, **139**, 1017–1023.

264. Bartlett, P.A., and Marlowe, C.K. (1987) Possible role for water dissociation in the slow binding of phosphorus-containing transition-state-analog inhibitors of thermolysin. *Biochemistry*, **26**, 8553–8561.

265. Fedosova, N.U., Champeil, P., and Esmann, M. (2002) Nucleotide Binding to Na,K-ATPase: The Role of Electrostatic Interactions . *Biochemistry*, **41**, 1267–1273.

266. Ledvina, P.S., Tsai, A.L., Wang, Z., et al. (1998) Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface: equilibrium, kinetic, and crystallographic studies. *Protein Sci.*, **7**, 2550–2559.

267. Radić, Z., Kirchhoff, P.D., Quinn, D.M., et al. (1997) Electrostatic influence on the kinetics of ligand binding to acetylcholinesterase. Distinctions between active center ligands and fasciculin. *J. Biol. Chem.*, **272**, 23265–23277.

268. Sinha, N., Mohan, S., Lipschultz, C.A., and Smith-Gill, S.J. (2002) Differences in Electrostatic Properties at Antibody–Antigen Binding Sites: Implications for Specificity and Cross-Reactivity. *Biophys. J.*, **83**, 2946–2968.

269. Molecular Operating Environment (MOE), Chemical Computing Group Inc., Montreal, QC, Canada.

270. Smart, O.S., Womack, T.O., Sharff, A., et al. (2011) *Grade (http://www.globalphasing.com)*, Global Phasing Ltd., Cambridge, UK.

271. Baell, J.B. (2013) Broad coverage of commercially available lead-like screening space with fewer than 350,000 compounds. *J. Chem. Inf. Model.*, **53**, 39–55.

272. Davis, B.J., and Erlanson, D.A. (2013) Learning from our mistakes: The `unknown knowns' in fragment screening. *Bioorg. Med. Chem. Lett.*, **23**, 2844–2852.

273. McGovern, S.L., Caselli, E., Grigorieff, N., and Shoichet, B.K. (2002) A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. *J. Med. Chem.*, **45**, 1712–1722.

274. Rishton, G.M. (2003) Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discov. Today*, **8**, 86–96.

275. McElhinny, C.J., Lewin, A.H., Mascarella, S.W., et al. (2012) Hydrolytic instability of the important orexin 1 receptor antagonist SB-334867: possible confounding effects on in vivo and in vitro studies. *Bioorg. Med. Chem. Lett.*, **22**, 6661–6664.

276. de la Figuera Gomez, T.H., Arques, J.S., Jones, R.A., et al. (1985) Pyrrole studies. Part 32. A novel ring-cleavage reaction of the pyridazine ring during the reaction of 6H-pyrrolo[3,4-d]pyridazines with dimethyl acetylenedicarboxylate. *J. Chem. Soc. Perkin Trans. 1*, 899–902.

277. Sheldrick, G.M. (2015) Crystal structure refinement with SHELXL. *Acta Crystallogr., Sect. C Struct. Chem.*, **71**, 3–8.

278. Eberson, L., Laerum, T., Undheim, K., et al. (1980) Studies on Anodic Substitution Reactions. XVI. Anodic Side-chain Substitution of 1,2,5-Trimethylpyrrole is Preceded by 2,5-Addition of Cyano Groups. *Acta Chem. Scand.*, **34b**, 747–751.

279. Karunaratne, V., and Dolphin, D. (1995) Oxidation with perhalogenated, water-soluble metalloporphyrins: application to oxidation of substituted 2-methylpyrroles. *J. Chem. Soc. Chem. Commun.*, 2105–2106.

280. Ohmiya, S., Noguchi, M., Ina, S., et al. (1992) Photoaddition Reaction of Pyrroles and Indoles to N-Methyl-2-pyridone. *Chem. Pharm. Bull.*, **40**, 854–857.

281. Sakurai, N., and Ohmiya, S. (1993) Photoaddition reaction of 1,2-dialkyl-indoles and -pyrroles to 1-methyl-2-pyridone via proton transfer from the 2-methylene group of the indole or pyrrole. *J. Chem. Soc. Chem. Commun.*, 297–298.

282. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria (2010) Gaussian 09, Revision C.01.

283. Schleyer, P. von R., Maerker, C., Dransfeld, A., et al. (1996) Nucleus-Independent Chemical Shifts: A Simple and Efficient Aromaticity Probe. *J. Am. Chem. Soc.*, **118**, 6317–6318.

284. E. D. Glendening, A. E. Reed, J. E. Carpenter, and F. Weinhold NBO Version 3.1.

285. Parr, R.G., Szentpály, L. v., and Liu, S. (1999) Electrophilicity Index. *J. Am. Chem. Soc.*, **121**, 1922–1924.

286. Koster, H., Craan, T., Brass, S., et al. (2011) A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.*, **54**, 7784–7796.

287. Smart, O.S., Womack, T.O., Sharff, A., Flensburg, C., Keller, P., Paciorek, W., Vonrhein, C., and Bricogne, G. (2011) Grade.

288. Bruno, I.J., Cole, J.C., Edgington, P.R., et al. (2002) New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallogr., Sect. B Struct. Sci.*, **58**, 389–397.

289. Macrae, C.F., Bruno, I.J., Chisholm, J.A., et al. (2008) Mercury CSD 2.0 -- new features for the visualization and investigation of crystal structures. *J. Appl. Crystallogr.*, **41**, 466–470.

290. R Development Core Team (2010) R: A language and environment for statistical computing.

291. Brown, H.C., Ayyangar, N.R., and Zweifel, G. (1964) Hydroboration. XVIII. The Reaction of Diisopinocampheylborane with Representative *cis* -Acyclic, Cyclic, and Bicyclic Olefins. A Convenient Systhesis of Optically Active Alcohols and Olefins of High Optical Purity and Established Configuration. *J. Am. Chem. Soc.*, **86**, 397–403.

292. Masamune, S., Kim, B.M., Petersen, J.S., et al. (1985) Organoboron compounds in organic synthesis. 1. Asymmetric hydroboration. *J. Am. Chem. Soc.*, **107**, 4549–4551.

293. Thomas, S.P., and Aggarwal, V.K. (2009) Asymmetric Hydroboration of 1,1-Disubstituted Alkenes. *Angew. Chemie Int. Ed.*, **48**, 1896–1898.

294. Zhu, S., and Buchwald, S.L. (2014) Enantioselective CuH-Catalyzed Anti-Markovnikov Hydroamination of 1,1-Disubstituted Alkenes. *J. Am. Chem. Soc.*, **136**, 15913–15916.

295. Homans, S.W. (2007) Water, water everywhere — except where it matters? *Drug Discov. Today*, **12**, 534–539.

296. Karunaratne, V., and Dolphin, D. (1996) Oxidation of substituted 2-methylpyrroles with perhalogenated metalloporphyrins: A one-pot synthesis of dipyrromethanes. *Tetrahedron Lett.*, **37**, 603–604.

297. Goldberg, R.N., Kishore, N., and Lennen, R.M. (2002) Thermodynamic Quantities for the Ionization Reactions of Buffers. *J. Phys. Chem. Ref. Data*, **31**, 231–370.

298. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D Biol. Crystallogr.*, **60**, 2256–2268.

299. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, W320–W324.

300. Biavardi, E., Favazza, M., Motta, A., et al. (2009) Molecular recognition on a cavitand-functionalized silicon surface. *J. Am. Chem. Soc.*, **131**, 7447–7455.

301. Gao, Y., He, X.-Y., Wang, Z.-B., et al. (2009) A new type of pseudorotaxanes based on cucurbit[6]uril and bis-cyanopyridyl alkane compounds. *Supramol. Chem.*, **21**, 699–706.

302. Cramer, J., Schiebel, J., Wulsdorf, T., et al. (2017) A False-Positive Screening Hit in Fragment-Based Lead Discovery: Watch out for the Red Herring. *Angew. Chemie Int. Ed.*, **56**, 1908–1913.

Acknowledgements

First and foremost I would like to thank my supervisor **Prof. Gerhard Klebe** for his scientific and personal guidance throughout the last four years. Without his support and invaluable experience in the world of science this work would not have been possible. I am also very grateful to **Prof. Martin Schlitzer** for reviewing this thesis and initially sparking my interest in organic and medicinal chemistry through his excellent lectures. For the participation in the thesis committee, I would also like to thank **Prof. Carsten Culmsee** and **Prof. Shuming Li**.

Many different people contributed to the publications that are included in this thesis. Most importantly, I would like to express my gratitude to **Stefan Krimmer** for the insightful discussions and productive cooperation. Through his tutelage, I was able to learn a lot about crystallography, ITC, and semi-obsessive desktop publishing. I am very grateful to **Prof. Andreas Heine** for generously sharing his vast knowledge of X-ray crystallography. I want to thank **Johannes Schiebel**, **Michael Betz**, **Tobias Wulsdorf**, **Veronica Fridh**, **Robert Karlsson**, **Uwe Linne**, **Nedyalka Radeva**, **Frederik Ehrmann**, **Nina Zitzer**, **Kristof Grohe**, **Eszter Najbauer**, and **Prof. Rasmus Linser** for the fruitful collaborations.

The value of a well organized laboratory and a healthy working atmosphere cannot be overstated. I am deeply grateful to **Hans-Dieter Gerber** for not only providing both of the above, but also a wealth of experience in organic chemistry and a constant disposition to share this knowledge. Many thanks also to **Rixa Kraut** and **Katharina Löbbert**, who contributed to the best lab experience I could imagine. I also want to thank **Lydia Hartleben** and **Christian Sohn** for their support in overcoming administrative and technical barriers. I would like to applaud the all members of the AG Klebe admin team for their great work. I am very thankful for the support of the analytical and technical departments of the institute. I want to thank **Nicole Bertoletti**, and **Christof Siefker** for the nice working atmosphere in the office (and also the never ending supply of pink unicorns in my direct line of sight). Many thanks to all the people of the Klebe, Kolb and Diederich labs, as well as honorary member **Marko Kljajic** and my good friends **Eva Schäfer** and **Nina Gerhards** for the great time on and off the job.

Finally, I want to give thanks to my family for their constant love and support. I am eternally grateful to my wife **Marlene Cramer** for her confidence, unquestioning loyalty, and faithful company for more than a third of my life.

Erklärung

Ich versichere, dass ich meine Dissertation

Inhibitor Synthesis and Biophysical Characterization of Protein–Ligand–Solvent Interactions

An Analysis of the Thermodynamics and Kinetics of Ligand Binding to Thermolysin

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen sind Zitate als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den . . .